PANEL DATA BASED FASHION SALES FORECASTING SYSTEMS

REN SHUYUN

Ph.D

The Hong Kong Polytechnic University

2016

The Hong Kong Polytechnic University

Institute of Textiles and Clothing

Panel Data Based Fashion Sales Forecasting Systems

Ren Shuyun

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

August 2015

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

--------------------------- (Signed)

--------------------------- (Name of student)

Abstract

Today's fashion industry is known to be information driven. The wise use of information for conducting sales forecasting helps a lot in enhancing the operations management of fashion companies. Numerous forecasting methods, such as statistical methods, Artificial Intelligence (AI) methods and many different kinds of hybrid methods, have been developed and studied in the literature for over two decades. However, good fashion sales forecasting is very difficult to achieve because the short product life-cycle and the ever-changing fashion trend make the demand of fashion product highly volatile.

In this thesis, to analyze and compare the cutting edge technologies of fashion sales forecasting, a comprehensive literature review was first conducted. It was found that pure one-dimensional time series based statistical methods were easy to implement and able to provide analytical solutions, but failed to yield a top forecasting performance. AI models were powerful for fashion sales forecasting, but they were computationally time consuming and usually required a large amount of historical data. Thus, in this thesis, panel data methods were introduced to conduct fashion sales forecasting. Panel data methods were widely used to do sales forecasting in various industrial settings. It could model the influence from other correlated products and some important related factors that pure time-series based

single-dimension statistical methods missed through its multi-dimensional data structure.

Useful fashion sales forecasting models should be applications oriented and hence industrialists' feelings and criteria should be examined. In order to reveal some insights regarding how Industrialists evaluated different major sales forecasting methods, an industrial survey was conducted with an aim to examine the industrialists' preference on fashion sales forecasting models. With the collected data, an analytic hierarchy process (AHP) analysis was conducted. To further investigate the preference of different decision makers with different roles, comparison studies were conducted by filtering the survey data into three category groups. Some important findings, including the usefulness of panel data based models, were obtained.

After that, in order to develop a versatile and innovative fashion sales forecasting application, a panel data based particle filter (PDPF) model was proposed for conducting fashion sales forecasting. The core advantage of this hybrid PDPF model was its three-dimensional correlation structure which could incorporate the influence of the previous sales of the specific product item, its price, and the effects brought by other correlated product items, into the forecasting model. A computational analysis, using real sales data, was conducted to further examine the forecasting performance of the proposed PDPF (versus other commonly seen methods reported in the literature).

It was found that the PDPF outperformed the other methods. Moreover, some important relationships, such as 1) the relationship between sales and corresponding price, 2) the relationship between the amount of historical data and the forecasting performance and 3) the relationship between the frequency of information updating and forecasting performance were all investigated. Important insights were generated.

Acknowledgements

This thesis was written based on the research work during my PhD study in the Institute of Textiles and Clothing, Faculty of Applied Science and Textiles, The Hong Kong Polytechnic University, from 2012 to 2015. Special thanks are given to The Hong Kong Polytechnic University for offering me the Research Studentship as the financial support for my PhD study. Coming to The Hong Kong Polytechnic University to pursue a PhD is a great experience to me. Over the past years of my PhD study, many people have given me strong supports to my research work and my life. I would therefore like to sincerely acknowledge them.

I would like to express my sincere thanks to my supervisor, Prof. Jason T.M. Choi, for giving me an opportunity to pursue the doctoral degree. Without for his kind advice and supports, it is impossible for me to finish this thesis research on time. His enthusiastic attitude towards research will always motivate me in the future endeavors. I would also like to sincerely thank Prof. K. K. Lai, Prof. Qi Xu, and Dr. Tracy Mok, the examiners of my PhD thesis research, for their many important advice and constructive comments on my thesis.

Besides, I would like to express my gratitude to Dr. Pui-Sze Chow, as well as my fellow colleagues during my PhD study: Dr. Chun-Hung Chiu, Dr. Jinhui Zheng, Dr. Na Liu, Dr. Bin Shen, Dr. Hau-Ling Chan and Ms. Wing-Yan Li. They shared their research experience and significantly contributed to my research through discussions and suggestions.

Finally, I would like to offer my deepest thanks to my dear parents, parents-in-law, and my husband Ivan Xuran LI for their constant love, supports and encouragements through years that cheered me up and let me successfully accomplish this challenging journey.

Table of Contents

Chapter 1 Introduction

1.1 Background of Study

In the fashion industry, a good sales forecasting tool for fashion products can help companies make better decisions. For example, fashion companies can reduce the amount of under-stocking or over-stocking in inventory control and increase the efficiency of production planning, and revenue management. Unfortunately, sales forecasting is a very difficult task in fashion retailing. This is partially a result of the high volatility of fashion demand, which is driven by the ever changing fashion trend; and partially a natural consequence brought by the lack of historical sales data owing to the short product selling season feature for most fashion products.

Fast fashion, a well-established business model, is an industrial trend. It is widely adopted by a number of international fashion companies such as Zara, Uniqlo, Mango and H&M in recent years (Choi 2013). In order to satisfy market demand, fast fashion companies tend to achieve a very fast response to fashion trend and customer's preference changing within a short lead time. As a result, fast fashion companies have to face the following conditions when conduct sales forecasting: (i) quick forecasting (Wang 2011) and (ii) using a limited amount of historical data to conduct prediction. Thus, fashion sales forecasting would become even more challenging under the fast fashion era.

In past decades, numerous forecasting methods were developed and studied in the literature, such as the statistical methods, the artificial intelligence (AI) methods and many kinds of hybrid methods. In fact, the statistical methods, such as auto-regression, exponential smoothing, ARIMA, SARIMA were probably the most widely used techniques for fashion sales forecasting. The advantages of these models are being fast, simple, well-explored, and easy to understand. However, the factors considered by these models are usually limited (Ni and Fan 2011). AI models were also employed to implement fashion sales forecasting (Au, et al. 2008; Frank, et al. 2003) and other forecasting for the fashion inudstry (Banica, et al. 2014). These methods are more versatile than statistical models. However, AI methods would require a substantial amount of time for conducting forecasting and the forecasting performance would largely depend on having sufficient historical data for training (Ma and Khorasani 2004). Combining the advantages of statistical methods and the AI methods, many kinds of hybrid methods were proposed in recent years (Aburto and Weber 2007; Choi, et al. 2011 and 2014; Kaya et al. 2014; Pai and Lin 2005; Thomassey, et al. 2005; Thomassey, et al. 2005; Vroman, et al. 1998). Many of these hybrid methods would perform well by employing the benefits of the component methods.

It was known that fashion products are quite different from many other non-fashion products. For instance, in fashion apparel, there are usually a lot of stock-keeping-units (SKUs) even under one single product line. In general, the demands of these SKUs are correlated. Thus, the sales of a specific fashion product would not only be influenced by the factors such as its size, color, price, etc, but also the sales of correlated items. This directly implies that the panel data based models could be good candidates for fashion sales forecasting.

The term "panel data" refers to a kind of data that contain time series observations of numerous of individuals. Therefore, observations in panel data involve at least two dimensions; a cross-sectional dimension and a time series dimension (that can be illustrated in Figure.1.1) (Hsiao 2003). Thus, the pure time series and pure cross-sectional data are special cases of panel data which are expressed in one dimension only (Baltagi 2008). Panel data usually contains observations of multiple phenomena obtained over a series of time periods for the same individuals. Panel historical data combining time-series data and cross-sectional data, would naturally provide more informative data for decision makers to conduct sales forecasting. Different from traditional time series models, the panel data model analyses each individual with multiple observations by including the unobservable effects which maybe correlated among all the individuals in the panel. Thus, it is able to model the influence from other correlated products and some important factors in which the other pure time-series data based methods would miss through its multi-dimensional data structure.

Motivated by the importance of fashion sales forecasting and the advantages of panel data models for conducting sales forecasting, this thesis investigates the panel data based forecasting methods for fashion sales and explores (i) the related literature, (ii) the strengths and weaknesses of a selected set of computational fashion forecasting models, including the panel data model; (iii) how industrialists evaluate these fashion sales forecasting models and their preference; (iv) the performance of a novel panel data based particle filter fashion sales forecasting model.

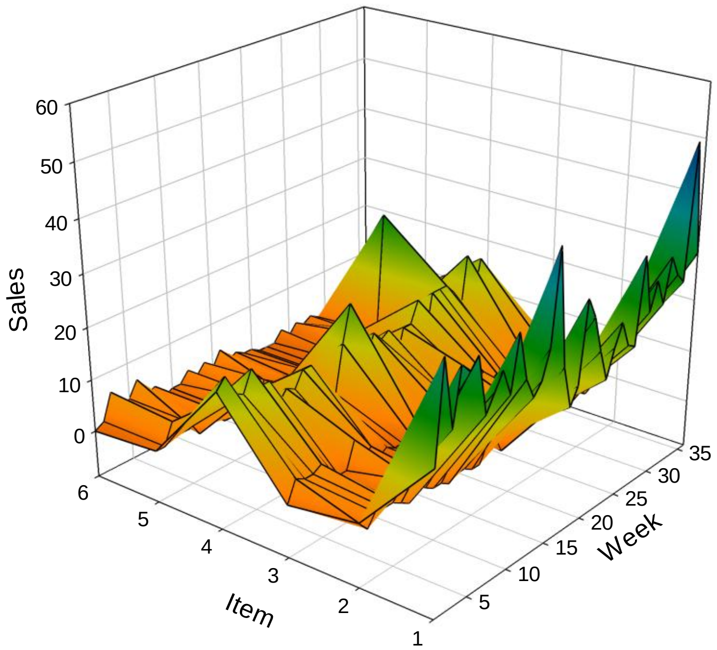


Figure 1.1 The panel data structure.

1.2 Research Objectives

The purpose of this research study is to develop panel data based methods for fashion sales forecasting and examine their performance compared with other popular methods. Moreover, this research goes one step further in exploring the industrialists' preference for choosing sales forecasting systems. To be specific, the four main objectives of this thesis research are listed as follows:

1. To analyze a set of computational models which can be applied for fashion sales forecasting and generate insights on their strengths and weaknesses.
2. To examine the industrialists' preference for different commonly seen sales forecasting systems.
3. To develop panel data based forecasting methods for fashion sales forecasting.
4. To examine the forecasting performance of these proposed methods and investigate the forecasting features of them.

1.3 Organization

This thesis consists of seven chapters and it is organized by the structure illustrated in Figure 1.2. To be specific, a comprehensive literature review on fashion sales forecasting methods, forecasting methods that can be used in fashion industry and applications of panel data method in sales forecasting, was conducted in Chapter

2. Next, the general analytical models, estimation methods and industrial applications of panel data sales forecasting models were reported in Chapter 3. A comparative study on panel data and other popularly used fashion sales forecasting methods was conducted with industrialists' inputs in Chapter 4. After that, a novel panel data based particle-filter (PDPF) model for fashion sales forecasting problem was studied in Chapter 5 and some further analysis on the forecasting features of the proposed PDPF model were shown in Chapter 6. Finally, this dissertation was concluded with discussions on future research in Chapter 7.

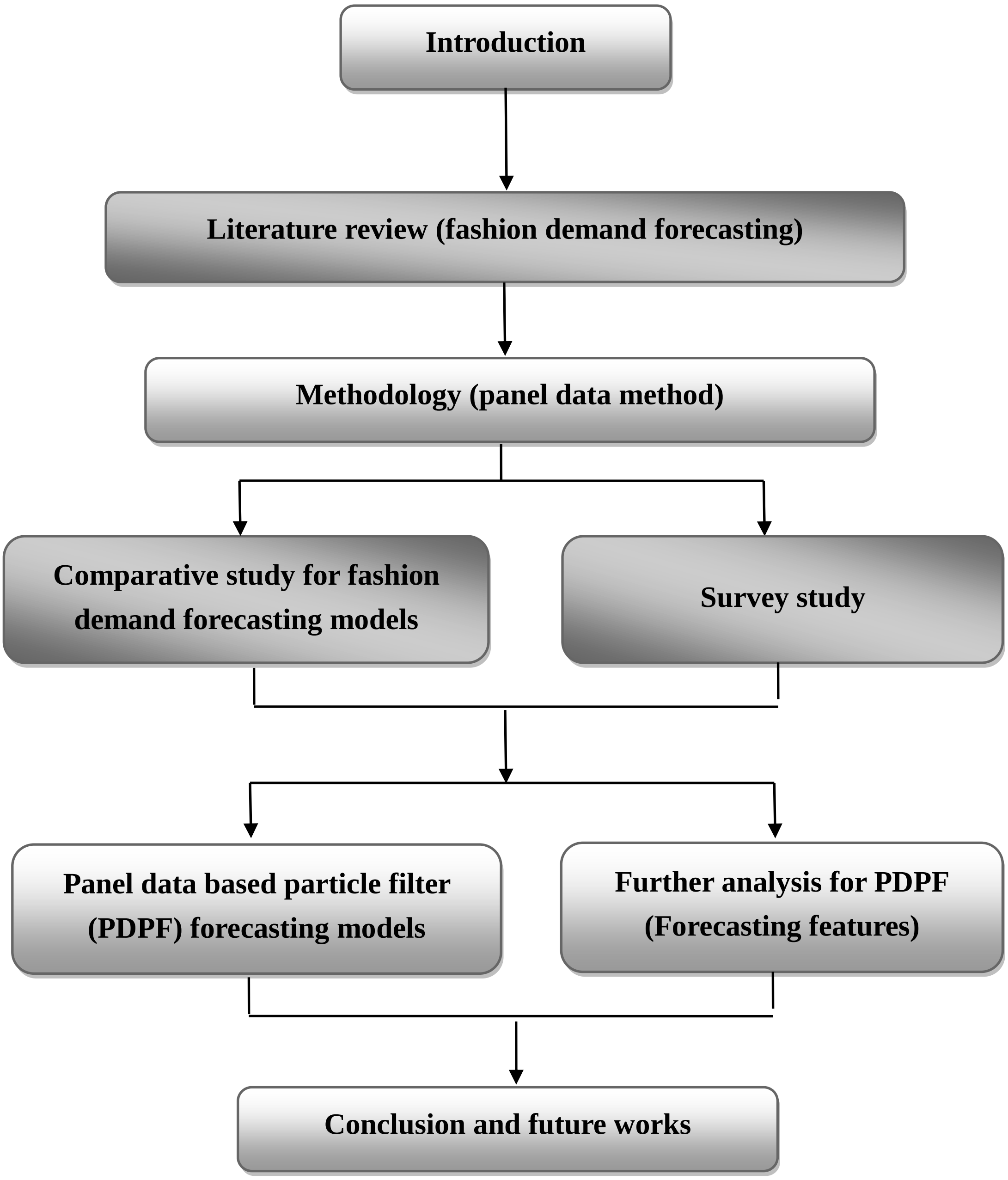


Figure 1.2 The organization of this thesis.

Chapter 2 Literature Review[1]

In the fashion industry, sales forecasting is a critical part of retail operations management. This chapter reviews the literature on fashion sales forecasting, and identifies the advantages and limitations of the major fashion sales forecasting methods.

2.1 Statistical Forecasting Methods

In conducting sales forecasting for fashion products, the classical approach starts by analyzing the product features (Nenni, et al. 2013). After that, fashion companies need to determine the forecasting approach. Traditional statistical methods, such as auto-regression, exponential smoothing, ARIMA, SARIMA are probably the most widely used techniques for fashion sales forecasting. The advantages of these models are being fast, simple, and easy to understand and implement. For the detailed applications of these methods for fashion sales forecasting, please refer to (Box and Jenkins 1976; Liu, et al. 2013; Thomassey 2014)). Another important advantage of these statistical methods is that they have a closed-form analytically tractable expression, which makes it easier to combine with other operations (e.g., inventory management) together. Recently, Yelland and Dong (2013) examined the applicability of a Bayesian forecasting model for conducting fashion sales forecasting, and they

[1] A part of this chapter was published as a journal paper in: Liu, N., S. Ren, T.-M. Choi, C.-L. Hui, and S.-F. Ng. 2013. Sales forecasting for fashion retailing service industry: a review, *Mathematical Problems in Engineering*, 2013. http://dx.doi.org/10.1155/2013/738675.

found that the Bayesian approach could yield better quantitative forecasting than many other methods. Despite being simple and fast, the statistic methods suffer some drawbacks for fashion sales forecasting: (i) The factors that may affect fashion sales forecasting considered by these models are usually limited (Ni and Fan 2011); (ii) it is difficult for simple single-dimensional statistical models to yield an excellent forecasting performance, in particular, compared with AI methods.

In recent years, with the emphasis on "big data" and the data driven knowledge-based operations 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 set of sample of individuals over time. It involves two dimensions: a time-series dimension T and a cross-sectional dimension N[2], and thus it provides two-dimensional observations on each individual in the sample. Panel data methods have become more and more important in the big data era (owing to the availability of data), even though the collection of panel data is more costly than the one-dimensional ones (Hsiao 2007). Panel data analysis has many advantages when comparing with the pure time-series single-dimensional econometric models (refer to Table 2.1). It usually gives a larger number of data points and incorporates much richer information from both time-series

[2] N: the number of cross-sectional units. Taking the sales data of apparel products as an example, N denotes the number of product categories.

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). The 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 sales forecasting problem when (i) the time series for all variables are shorter, and (ii) cross-sectional information on these variables is available. Obviously, these situations are satisfied for many fashion retail operations.

Table 2.1 Advantages of panel data based models compared to time-series based models

	Panel data	**One-dimensional data**
Sample data	Two dimensions(N,T)	Only one-dimensional 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

2.2 AI Forecasting Methods

With the advance of computing technologies, artificial intelligence (AI) models, which are more versatile than statistical models, have been widely employed to implement fashion sales forecasting (see Frank, et al. (2003), Au et al. (2008), El-Bakry and Mastorakis (2008) and Yu et al. (2012)). Artificial neural network (ANN) methods are a popular set of forecasting models for predicting fashion product's sales. It is well known that even a simple ANN would take a substantial amount of time to complete a basic sales forecasting task (e.g., it may take several minutes, and evolutionary neural networks (ENN) would even take a longer time (Au, et al. 2008)). The long computational time becomes a major hurdle for the deployment of many ANN and ENN based forecasting models in real world fashion sales forecasting. ANN models have been proven that they are able to provide satisfactory results (in terms of forecasting accuracy) in different forecasting domains (Olson and Mossman 2003; Yoo 1999; Zampighi et al. 2004). Despite yielding high forecasting accuracy, ANN and ENN based forecasting models are very time consuming to run due to their utilization of the gradient-based learning algorithms. To overcome this drawback, the extreme learning machine (ELM) (Pao et al. 1994) based models were proposed in recent years for sales forecasting in fashion by Sun et al. (2008). ELM's performance in sales forecasting is proven to be better than many back propagation neural networks based methods (Zhu et al. 2005; Huang et al. 2006). After that, an extended ELM method (EELM) based algorithm was proposed to further enhance the forecasting performance by Yu et al. (2012). Although being more stable than ELM, EELM still needs a substantial amount of time to conduct prediction. Besides, another popular forecasting method for predicting fashion product's demand is the grey method (GM), that has been known to be a very efficient method to deal with time-series demand forecasting problems with insufficient historical data (Mostard et al. 2011; Mengi and Altas 2011; Hui et al. 2005; Zhu et al. 2005; Sun et al. 2008; homassey et al. 2005; homassey et al. 2005; Aksoy et al. 2012). Fuzzy logic based models are able to better identify nonlinear relationships in the input data, that makes them conduct good performance for fashion demand forecasting (Liu et al. 2014). For fashion forecasting, the related reviews can be found in (Liu et al. 2014), Nenni et al. (2013), Thomassey Thomassey (2014), and Choi et al. (2011).

Despite being powerful, AI methods usually require a substantial amount of time for conducting forecasting and the forecasting performance largely depends on having sufficient historical data for training.

The limitations of both AI methods and statistical methods hence call for the development of innovative new methods. For example, a series of hybrid models, which combine the advantages of statistical methods and the AI methods, were proposed in the recent literature (Aburto and Weber 2007; Choi, et al. 2011; Pai and Lin 2005; Thomassey, et al. 2005; Thomassey, et al. 2005; Vroman, et al. 1998; Wong and Guo 2010) and they would be reviewed below.

2.3 Hybrid Forecasting Methods

Combining advantages of statistical methods and the AI methods, a series of hybrid models were proposed in the literature of fashion sales forecasting in recent years. As such, many of them were considered to be more efficient than the pure statistical models and pure AI models. Hybrid methods employed in the fashion forecasting literature often combine different schemes such as the fuzzy model, ANN, ELM with other techniques such as statistical models, and the grey model (GM), etc. In the following, several examples are discussed.

2.3.1 Fuzzy Logic Based Hybrid Methods

After Vroman et al. (1998) developed a fuzzy-adaptive model to predict the sales for new items without historical data, several fuzzy logic based hybrid models were proposed in the literature to conduct fashion sales forecasting. Forecasting results showed that the proposed fuzzy-adaptive model outperformed the conventional Holt-Winter method. After that, Thomassey et al. (2005a) used the fuzzy logic concept to perform fashion forecasting. Their new model allowed an automatic learning of the non-linear explanatory variables' influence. Notice that their model required a subjective expert judgment for the learning process which posed a challenge for its real world application in the fashion retailing industry. Extending Thomassey, et al. (2005a), Thomassey et al. (2005b) proposed a forecasting system which was composed of a mid-term forecasting model called AHFCCX and a short term predict model called SAMANFIS. Their proposed model was based on multiple models such as fuzzy logic, neural networks and evolutionary procedures. AHFCCX could realize a mean term forecasting based on historical data of precedent historic season and the SAMANN model would then use this mean term forecasting and last weeks' sales to obtain a short term prediction. The authors argued that their proposed method was versatile in processing the uncertain data. Recently, Yesil et al. (2012) applied a hybrid fuzzy model to conduct fashion forecasting. To be specific, they combined the fuzzy logic model and the statistical model to conduct forecasting. In their hybrid method, they calculated a final forecast for weekly demand based on the weighted average of forecasts that were generated by multiple methods. They argued that their proposed method could achieve very high accuracy.

2.3.2 Neural Network Based Hybrid Methods

In neural network (NN) hybrid models, Vroman et al. (2001) employed a NN model with the "corrective coefficients of seasonality" for a mean-term forecasting horizon. They argued that their proposed hybrid method can also conduct forecasting for short and discontinuous time series. They reported significantly good results with their proposed NN hybrid model and believed that the outstanding performance came from the NN's ability of mapping the nonlinear relation between data inputs and output. Thomassey and Happiette (2007) developed a hybrid neural clustering and classification scheme for conducting sales forecasting of new apparel items. Their model could increase the accuracy of mid-term forecasting in comparison with the mean sales profile predictor. Notice that ANN based models could also be combined with other techniques like grey method (GM), and auto-regressive models. For instance, a two-stage dynamic forecasting model, which contained neural networks and the auto regressive technique, was applied for fashion retail forecasting in Ni and Fan (2011). In their model, Ni and Fan used neural networks to establish a multi-variable error forecasting model. Their model developed the concept of 'influence factors' and divided the 'impact factors' into two distinct stages (long term and short term). Their computational experiment showed that the multi-variable error forecasting model can yield good prediction results for fashion retail sales forecasting problems. Aksoy et al. (2012) combined the fuzzy method and neural networks to form a new system called the adaptive-network based fuzzy inference system. Their proposed new system combined the advantages of both systems, namely the learning capability of neural networks and the generalization capability of fuzzy logics. Rather recently, Choi et al. (2012) applied an ANN and GM based hybrid model for fashion sales forecasting with respect to color. They systematically compared ANN, GM, markov regime switching, and GM+ANN hybrid models. They revealed that the

GM(1,1) and ANN hybrid model were the best ones for forecasting fashion sales by colors in the presence of very few historical data.

2.3.3 ELM Based Hybrid Methods

The extreme learning machine (ELM) is quick in conducting forecasting (Yu, et al. 2011). Despite the fact that it is not perfect because of its unstable nature, its “fast speed” makes it a very good candidate to be a component model for the more advanced hybrid model for fashion forecasting. For example, to investigate fashion sales series from a new perspective, Wong and Guo (2010) proposed a hybrid intelligent model combining the learning algorithm-based neural network and the heuristic fine-tuning process. Novel learning algorithm-based neural network was used to generate initial sales forecast and heuristic fine-tuning process was adopted to obtain more accurate final sales forecast. They claimed that the performance of their proposed model was superior to the traditional ARIMA models and two recently developed neural network models and suggested that the patterns in its same-period time series were much simpler than its original pattern if the original monthly time series exhibited a strong level of seasonality. Xia et al. (2012) examined a forecasting model based on ELM with the adaptive metrics. In their model, the inputs would solve the problems of amplitude changing and trend determination, which in turn helped to reduce the effect of the over fitting of networks. Yu et al. (2012) used ELM and Grey Relational Analysis (GRA) to develop a fashion color forecasting hybrid method. Their computational result with real empirical data illustrated that their proposed model would outperform several other competing models in forecasting fashion color.

2.3.4 Other Hybrid Methods

In addition to the types of hybrid methods reviewed above, some other innovative forecasting combined methods were also reported in the literature. For example, Choi et al (2011) employed a novel hybrid SARIMA wavelet transform (SW) method for fashion sales forecasting. Using both real data and artificial data sets, they showed that with a relatively weak level of seasonality and a highly variable seasonality factor, their proposed SW method would outperform the classical statistical methods. They concluded to say that the SW method was suitable for conducting volatile retail sales forecasting in fashion. Thomassey and Happiette (2007) developed a hybrid method which was based on an existing clustering technique and a decision tree classifier. Their proposed hybrid method was known to be very useful for estimating the sales profiles of new items in fashion retailing in the absence of historical sales data. Ni and Fan (2011) established a combined method which included the auto regression and decision tree method (called the ART method). They proposed that their hybrid method performed very well for fashion sales forecasting. Table 2.2 summarized the reviewed hybrid methods.

Table.2.2 The summary of hybrid methods-based fashion retail sales forecasting models

Method		Paper	Domain	Findings
Fuzzy	Holt Winter	(Vroman, et al. 1998)	New items	The proposed fuzzy-adaptive model would control the weight factors of an exponential-smoothing forecasting method, and it could be applied for new item sales forecasting problems.
	CCX	(Thomassey, et al. 2002)	Mean term	The method used fuzzy logic abilities to map the non-linear influences of explanatory variables to conduct sales forecasting. However, expert judgment was required for the learning process for this method.
		(Thomassey, et al. 2002)	Mean term	This method allowed a fashion company to obtain mean term forecasting to pass commands to providers.
	NN	(Thomassey, et al. 2005)	Short-term	The method could perform short-term forecasting by re-adjusting mean-term model forecasts with the inputs of real sales.
	Distribution of Aggregated forecast and Classification	(Thomassey, et al. 2005)	New items Insufficient data	Their proposed items forecasting model would estimate the items sales of the same family without requiring any historical data.
	NN	(Aksoy, et al. 2012)	Short term	The model greatly improved the accuracy of forecasting results for the short horizon of one month.
	NN	(Yesil, et al. 2012)	Fast forecasting	Using fuzzy logics, the combiner would calculate a final forecast for each week's demand as a weighted average of forecasts that were generated by different

				methods. This combined forecast would achieve better accuracy than any of the individual forecasts.
ANN	CCX	(Vroman, et al. 2001)	Mean term	Considering noisy data and multiple explanatory variables (controlled, available or not) related to the sales pattern, the proposed model performed well.
	Classification	(Thomassey and Happiette 2007)	New items	Neural clustering and classification model globally increased the accuracy of mid-term forecasting in comparison with the mean sales profile predictor.
	ELM+ Harmony search	(Wong and Guo 2010)	Mean term	The learning algorithm integrated with an improved harmony search algorithm and an extreme learning machine could improve the network generalization performance and was better than traditional ARIMA models and two recently developed neural network models in fashion sales forecasting.
	ART	(Ni and Fan 2011)	Two stages: Long term and short term	Combining the ART model and the error forecasting model based on neural networks, an adjustment improving model which could be applied for the fashion retail forecasting was developed.
	GM	(Choi, et al. 2012)	color trend insufficient data	GM+ANN hybrid models were examined in the domain of color trend forecasting with a limited amount of historical data. The GM + ANN hybrid model was the best one for forecasting fashion sales by colors in which only very few historical data was available.
ELM	Statistic	(Yu, et al. 2011)	Fast forecasting	A comparison with other traditional methods showed that the ELM sales forecasting model is quick and effective.
	Metrics	(Xia, et al. 2012)	Sufficient data	The adaptive metrics of inputs could help solve the problems of amplitude changing and trend determination, and reduce the effect of the over fitting of the neural networks. The model

				outperformed auto-regression (AR), ANN and ELM models.
	GRA	(Baltagi and Griffin 1997)	Color trend	With real data analysis, the results showed that the ANN family models, especially for ELM with GRA, outperformed the other models for forecasting fashion color trend.
SARIMA	Wavelet	(Yu, et al. 2012)	highly volatile sales	For real data with a relatively weak level of seasonality and a highly variable seasonality level, the SW hybrid model performed well.
Decision tree	Clustering	(Thomassey and Fiordaliso 2006)	Mean term	The proposed model, based on an existing clustering technique and the decision tree classifier, was useful to estimate sales profiles of new items in the absence of historical sales data.
	Auto-regressive technique	(Ni and Fan 2011)	Short term	Combining the ART model and the error forecasting model based on neural networks, the “adjustment improving” model was applicable for the fashion retail forecasting.

Chapter 3 Panel Data Models[3]

3.1 Introduction

In Chapter 2, various models for fashion sales forecasting were examined. It was found that the panel data model could be a very useful one for fashion sales forecasting. In this chapter, several important and commonly examined panel data based models were reviewed and insights regarding their strengthens, weaknesses, and applications were generated.

Over the past decades, panel data models and forecasting analysis were used in many research areas. Baltagi (2008) gave a pioneering literature review of panel data based forecasting. He found that panel data estimators performed very well in forecasting, although the forecasting performance of various panel data estimators vary from one empirical example to another.

As reviewed in Chapter 2, sales forecasting methods based on time-series data were well-established. Under these methods, the sale of each fashion item was estimated independently, and there was a lack of consideration on the unobservable correlation of individuals. This would make the pure time-series approaches far from effective. Panel data models (Getis 2007), with a cross-sectional dimension I and a time-series dimension T, could provide the possibility of learning an individual

[3] A part of this chapter was prepared as a review paper under the journal review.

item’s sales pattern by observing the pattern of others, in addition to the information on that particular individual item’s pattern. In fact, a panel data structure is one that follows a set of individual sample over time, and it thus observes each individual in the sample together. (Hsiao 2003). Combining cross-section and time-series data is able to better handle the unobservable effects (individual-specific) which might have serious correlations with other explanatory variables. As described by Jang and Shin (2014), panel data techniques were proposed by Airy in 1861 for the astronomical data analysis. Since the 1960s, the panel data based methods have become widely available for forecasting labor economics all around the worlds. Two of the most classical panel data sets, the University of Michigan’s Panel Study of Income Dynamics (PSID) and the National Longitudinal Surveys of Labor Market Experience (NLS) , were also constructed in the U.S. in 1968 (Hsiao 2003). Borus (1981) and Juster (2000) classified and discussed other mainstream panel data sets and found that economists were of interest to analysis data in panel structures. Panel data sets for economic research outperformed either conventional cross-sectional or time-series data sets in several major aspects (e.g., Hsiao (1985a, 1995, 2000)) (refer to Table 3.1). Panel data models would usually have the advantages compared with conventional one dimensional data model, such as: larger degrees of freedom, more data points and smaller collinearity among explanatory variables. Therefore the efficiency of econometric estimation would be enhanced by using panel data structure (Hsiao 2003).

Table 3.1 Advantages of panel data methods compared to pure single-dimensional time-series methods

	PANEL DATA	ONE-DIMENSIONAL DATA
Sample data	Involves two dimensions(N,T)	Involves 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

First, Telser (1962) used pooled data to do the demand analysis for branded goods. In his analysis, the customer purchase behaviors were not only investigated by the price of the studied product, but also the price of other branded products. Baltagi (2008) gave a detailed review of forecasting applications with panel data, and suggested that panel data technique performed very well in forecast probably due to their simplicity. However, he also pointed out that the forecasting performance of different panel data estimators might vary from one empirical example to another. Later on, Babel, et al. (2008) proposed a stochastic mortality model to capture both common age effect and common time effect of German mortality rates using panel data structure. Their study suggested that the panel structure model would allow a direct interpretation of the parameters and lead to promising forecasting results. Gholami et al. (2009) conducted a comprehensive study on the patterns and mechanisms of “spillovers”, an international information and communication technology, by constructing a 37 countries’ panel data from 1996 to 2004. Issler and Lima (2009) proposed a novel approach to econometric forecast of stationary time series within a panel-data framework and showed that the number of forecasts and the number of time periods would increase without bounds using panel data sequential asymptotic. Furthermore, the empirical studies of the exchange rate model with panel data revealed that there are potentials in forecasting accuracy increase, estimation precision improvement, and wider application capability when using panel data. In contrast to single-section models, panel models were often able to outperform a random walk in out-of sample tests (see (Hulley and McWalter 2008)). Baltagi et al. (2012) proved that estimators that ignored heterogeneity/spatial correlation might perform poorly in RMSE forecasts under their Monte Carlo studies. Recently, combining the pane cointegration model and the particle filter method, Li et al. (2013) investigated an energy price forecasting model for several interconnected regions based on a two-stage panel data. Their empirical results indicated that their proposed model performed better than some competing AI methods.

In terms of sales forecasting, some related studies were reviewed in the following. Levi (1986) estimated the dynamic demand for cigarette consumption based on panel data structure. The panel is composed of pooled time-series over the period from 1963 to 1980 and cross-section data of 46 states. In the study of dynamic demand for gasoline over the period 1960–1990 across 18 OECD countries, Baltagi and Griffin (1997) proposed a panel-data based technique to estimate reliable price and income elasticities by pooling the data. Then, Baltagi et al. (2012) reconsidered the two US panel datasets on residential electricity and natural-gas demand and compared the out-of-sample forecast performance. The results once again proved that pooled data would offer better out-of-sample forecast. After that, Kesavan et al. (2010) constructed a simultaneous equations model using the panel data to provide joint forecasts of annual cost of goods sold, inventory, and gross margin for retailers. They claimed that the sales forecasts from their proposed panel data model were more accurate than the consensus forecasts from equity analysts.

Undoubtedly, the panel data based methods were well-applied in practice. In the following, important details of panel data based models would be examined.

3.2 Analytical Models

A common classification scheme of several panel data based models that were used in forecasting is shown in Figure 3.1. Notice that from the stage dependency perspective, panel data based models can be classified as static panel data models and dynamic panel data models. The static model is more structural than behavioral while the dynamic model gives a representation of the behavior of the system's static components. Considering the impact of the individual-specific effects, a panel data model can be classified as the fixed-effects model and the 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. Moreover, based on the dependence relationship in error terms, panel data models can be categorized as spatial correlation models and serial correlation models. Spatial correlation models 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. In the following, more analytical details of these models would be introduced.

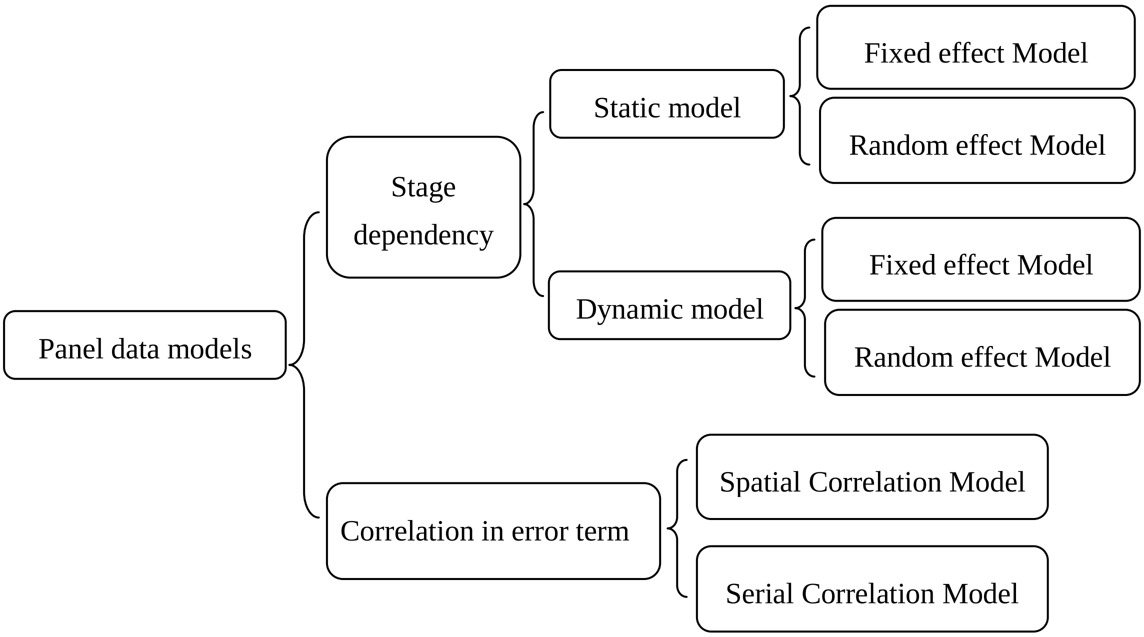


Figure 3.1 The proposed classification of panel data model.

3.2.1 Static Models

Following (Hsiao 2003), the common panel data regression model is presented as:

$$y_{it} = \alpha + \beta^{'} X_{it} + u_{it} \qquad (1)$$

where $i = 1, 2, ..., N$ denotes N individuals. $t = 1, 2, ..., \mathrm{T}$ 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). The basic assumption of such models is that, conditional on the observed explanatory variables, the effects of all omitted (or excluded) variables are driven by three types of variables:

1. *Individual time-invariant:* The individual time-invariant variables have two characteristics: (i) vary across cross-section (ii) are the same for a given cross-section over time.
2. *Period individual-invariant:* The period individual-invariant variables have two characteristics: (i) vary over time (ii) are the same for all cross-sectional units at a given time.
3. *Individual time-varying:* The individual time-invariant variables have two characteristics: (i) vary across cross-section (ii) are the same for a given cross-section over time.

Next, the "fixed-effects" and "random-effects" models would be examined.

A. Fixed-effects Model

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

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

Assuming there are no time-specific effects, and only individual-specific effects exist 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 was studied by Wallace and Hussain (1969) and Swamy and Arora (1972). The advantage of fixed-effects inference is that there is no need to make a 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 estimates of α_i and β would be obtained by minimizing $S = \sum_{i=1}^{N} U_i^{'} U_i = \sum_{i=1}^{N} (y_i - e\alpha_i - X_i\beta)^{'} (y_i - e\alpha_i - X_i\beta)$ (3).

Taking partial derivatives of S with respect to α_i and setting them equal to zero yields

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

where

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

Then, the estimation of β could be obtained as follows:

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

Observe the OLS estimator is a consistent estimator (Amemiya 1985) under the fixed-effects assumption when T tends to infinity. Another finding is that no correlation between the error term and any of the explanatory variables is a necessary condition for having unbiased and consistent parameter estimation under OLS.

B. Random-effects Model

Unlike the fixed-effect model in which it treats the effects of omitted individual-specific variables α_i as fixed constants over time, the random-effects model treats the individual-specific effects as random variables. The advantage of random-effects inference is that the estimation methods can derive the result efficiently because the number of parameters is fixed. The disadvantage is that specific assumptions needed to be made by the decision maker about whether the pattern of correlation exists between the effects and the explanatory variables (Hsiao 2003). The linear regression, which is called two-way component model in Baltagi (2008), can be written in 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 (2008) showed 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:

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

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

where

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

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

Note that Taylor (1980) compared the within estimator with the GLS estimator for the random-effects one-way component model by using a finite sample and found that the feasible GLS is more efficient than the 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 MLE estimator of the slope coefficients. The CV estimator is consistent for the static model no matter whether the effects are fixed or random.

C. Comparisons of Fixed-effects and Random-effects Models

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 model means that the individual-specific effects from other products or time period fluctuate following a distribution. If the effect in a panel data model is modeled as being random, the features of individual behavior would be learned from the features of other observed individual behaviors, rather than about these particular units themselves. It is important to test and estimate the variance of these random-effects across different products. While under the fixed-effects assumption, the effect from other product or time period is the average effect of each fashion product, expressed by the regression coefficient. Hsiao (2003), Baltagi (2008) and Lee and Yu (2010) all gave a detailed discussion on the choice between random-effects and fixed-effects models. The comparison of these two kinds of effects is illustrated in Table 3.2. In practice, the selection of the appropriate model is important to ensure that the various statistics are estimated correctly and properly (Borenstein, et al. 2010). Observe that Hausman (1978) showed that using a fixed-effects model would produce significantly different results from a random-effects model in his estimation of a wage equation using a sample of 629 high school graduates followed over a six years period. Hsiao (2003) found that whether to treat the effects as fixed or random would make no significant difference when T is large. For the choice between fixed-effects and random-effects, Hsiao (2003) gave numerous examples in which the purpose of analysis would determine how to choose them.

Table 3.2 Comparisons between fixed-effects and random-effects models

	Fixed-effects	Random-effects
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 period and all individuals	Estimate the mean of the true effects distribution for all time period and all individuals

3.2.2 Dynamic Models

Panel data based models were widely used to estimate the parameters of dynamic econometric models. In fact, dynamic panel data models which were able to describe the dynamic relationship between explained variable and explanatory variables were widely used to deal with sales forecasting problems in various research areas, such as energy consumption, tourism demand, water demand, etc. In the following, some commonly used dynamic models and the respective estimation methods would be introduced. Estimation here means using the demand historical data to estimate the unknown parameters of the panel data model.

A. Common Regression Model

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

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

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.

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

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

Dynamic models are widely used in the demand forecasting applications, including tourism demand forecasting, consumer products demand forecasting, electricity and natural-gas consumption and etc. Allowing for 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.

B. Estimation for Dynamic Models

If the model contains exogenous variables in addition to the lagged dependent variables, the situation becomes different. Besides the covariance estimator and the MLE may be inconsistent, the interpretation of the estimation may also depend on the initial condition assumption (Anderson and Hsiao 1982). Anderson and Hsiao (1981) studied the problem of estimating a dynamic error components model when either the number of cross-sectional unit *N* or the number of time point *T* tends to infinity.

Different assumptions about the initial conditions and the derivative models were also examined. Their study showed that when T tends to infinity, the MLE is consistent under all assumptions on different initial conditions. When N tends to infinity and T is fixed, whether the MLE is consistent would depend on assumptions about the initial condition. Anderson and Hsiao (1982) showed the sensitivity of MLE estimators by alternating the assumptions about initial conditions (the assumption for α_i, λ_t and u_{it}.) and asymptotic plans. They argued that the advantage of these estimators is their consistency (irrespective of the initial conditions and whether T or N or both tend to infinity). The generalized method of moments (GMM) estimator, developed by Hansen (1982), provides a convenient framework for dynamic models. Bond (2002) reviewed the use of GMM estimators in the model which contained endogenous or predetermined explanatory variables, with a large number of cross-section units observations for a small number of time periods ($N \rightarrow \infty$ for fixed T) and suggested that GMM was useful and efficient for the estimation of this kind of panel data model. Arellano and Bond (1991) applied the one-step GMM estimation method to estimate an unbalanced panel data consisting of OECD countries over 1978 to 1999 in (Liu 2004). A simple partial adjustment model was then used to specify the energy demand. From the empirical results, the one-step GMM estimator outperformed the "within estimator" and the OLS in terms of sign and magnitude. Hsiao (2003) revealed 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) supported this conclusion by Monte Carlo simulation studies.

For the fixed-effects model: If lagged dependent variables appear as explanatory variables, strict exogeneity of the regressors no longer holds. The MLE or the covariance 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 proper formulation of initial conditions. Hsiao et al. (2002) suggested a transformed MLE and a computationally simpler minimum distance estimator (MDE) for fixed-effects formulation and conducted 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 showed that IV and GMM estimators both do not need the formulation of initial conditions. Furthermore, the likelihood approach appears to dominate the GMM approach 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 ordinary-least-squares (OLS) estimator for dynamic models is biased and inconsistent (Hsiao 2003). For random effect 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. Prior studies, such as Nerlove (1971), Sargan and Bhargava (1983), and Nerlove and Balestra (1996), discussed the ML estimation of the dynamic random effects model. The MLE, the IV, and the GMM estimators were proven to be consistent (Hsiao 2003), although OLS estimator was 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 (e.g, the MLE, the CV, the IV estimators, and the generalized least-squares estimator (GLS)). Anderson and Hsiao (1981) studied the problems of estimating a dynamic model with error components under assumptions and Table 3.3 lists consistency properties of different estimators under different assumptions about the initial observations and the way T and N tend to infinity. From Table 3.3, it can be easily found that the consistency of MLE, CV and GLS depends on the assumption of initial observations and the way T and N tend to infinity (except for the IV estimator). Alvarez and Arellano (2003) suggested that MLE is more efficient and robust than GMM when both T and N tend to infinity.

Table 3.3 Statistical properties of different estimators under different assumptions about the initial observations and the way T and N tend to infinity

	Interpretation of the model	**Statistical properties**			
		MLE	CV	GLS	IV
case 1: y_{i0} is fixed	$T \to \infty$, $N \to$ fixed	Consistent	Consistent	Consistent	Consistent
	$T \to$ fixed, $N \to \infty$	Consistent	Inconsistent	Consistent	Consistent
case 2: y_{i0} is random with a common mean	$T \to \infty$, $N \to$ fixed	Consistent	Consistent	Consistent	Consistent
	$T \to$ fixed, $N \to \infty$	Consistent	Inconsistent	Consistent	Consistent
case 3: w_{i0} is fixed	$T \to \infty$, $N \to$ fixed	Consistent	Consistent	Consistent	Consistent
	$T \to$ fixed, $N \to \infty$	Inconsistent	Inconsistent	Consistent	Consistent
case 4: w_{i0} is random with a stationary distribution	$T \to \infty$, $N \to$ fixed	Consistent	Consistent	Consistent	Consistent
	$T \to$ fixed, $N \to \infty$	Consistent	Inconsistent	Consistent	Consistent

Summary: Dynamic panel data based models containing lagged dependent variables would allow us to better understand the dynamics of adjustment. However, when lagged dependent variables also appear as explanatory variables, the estimator for static model may no long be consistent and efficient. The statistical properties of some common estimators for dynamic models are summarized in Table 3.4. Although there are many kinds of theoretical estimators for dynamic panel data, the estimation performance is different from situation to situation in practice. The OLS methods and several different methods for estimating parameters in the presence of lagged endogenous variables were discussed in Balestra and Nerlove (1966). Nerlove (1967) conducted Monte Carlo studies and suggested that the OLS method would overestimate when N or T or both tend to infinity. Then, Houthakker et al. (1974) presented a variance component technique developed by Balestra and Nerlove (1966) for estimating the dynamic model and suggested that the variance component technique could provide very satisfactory results, while OLS and IV could not estimate well for the demand of gasoline and residential electricity case. After that, Babel et al. (2008) used an OLS estimator to estimate a time-dynamic stochastic model by utilizing a panel data approach for German mortality forecasting and obtained satisfied estimation results. Garín-Muñoz and Montero-Martín (2007) employed the GMM-DIFF estimator proposed by Arellano and Bond (1991) to estimate a panel data model which includes lag dependent explanatory variables and yielded satisfactorily good performance model. This estimator are suitable for the case that the dependent variable which lagged two periods or more, because they were valid instruments for the lagged dependent variable. The parameters therefore would achieve a consistent and efficient result via estimation.

Table 3.4 Statistical properties of some common estimators for dynamic models

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			

3.2.3 Spatial Correlation Model

The term spatial correlation model, first used in 1967, included critical ideas such as distance-decay and spatial interaction (Getis 2007). Good examples of the spatial correlation model could be found in the econometric analysis of economic panel data in which many regional effects were modeled by spatial correlation models (Jang and Shin 2014). The spatial econometrics literature was usually supported by studies based on the dependence among observations across space and used the so-called spatial weights matrix W to describe the spatial arrangement of the geographical units in the sample (Baltagi 2008). Since the studies by Anselin (1988), the spatial panel data models have been increasingly attractive in empirical economic research. Spatial panel data models with spatial error autocorrelation, including a spatially lagged dependent variable, also received much more attention in the regional science literature (Sarkar 2003). The general panel data models would allow us to control for heterogeneity across multiple individual units (Baltagi 2008), while spatial panel data models could control both heterogeneity and spatial correlation (Baltagi, et al. 2003). If the spatial dependence between observations is specified, the spatial panel data model may incorporate a spatial autoregressive process in the error term (which is termed as the spatial error model), or contain a spatially autoregressive dependent variable (which is known as the spatial lag model). In this aspect, the traditional panel data only captures the "average" or representative behavior in the cross-section diminution. It results in average effects across spatial units that would lead to missing of the differences in behaviors among individual spatial units (Quah 1996). Panel data models allow researchers to control heterogeneity across spatial units, while spatial panel models allow researchers to control both heterogeneity and spatial correlation (see Baltagi (2008) for more discussions). Let W denote a ($N \times N$) spatial weight matrix describing the spatial arrangement of the spatial units, w_{ij} denote the (i, j)th element of W, where i and $j = 1, \ldots N$. The traditional spatial error autocorrelation model could then be written as (Baltagi 2008):

$$Y_{it} = \beta^{'} X_{it} + u_{it}, \qquad (11)$$

The disturbance vector form is given by:

$$U_t = u + \phi_t, \qquad (12)$$

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

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

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

The spatial econometric literature (LeSage and Pace 2010) indicated that the OLS estimator of the response parameters was unbiased for spatial error autocorrelation model, but it lost the efficiency property. For the case when the specification contained a spatially lagged dependent variable, the OLS estimator of the response parameters not only lost the property of being unbiased but was also inconsistent (Elhorst 2003). To overcome this problem, (Anselin 1988) and (Anselin and Hudak 1992) used maximum likelihood techniques to conduct the estimation. Following that, GMM estimator was proven to be robust to spatial dependence among the error terms in spatial cross-section models (in (Conley 1999); (Anselin 1999)). Yu et al. (2008) established and discussed the 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) extended the previous studies and examined the performance of QML, 2SLS and GMM estimations for the unstable cases where there are unit roots generated by temporal and spatial correlations. They suggested that, the QML estimation's consistency requires T tending to infinity, while the GMM is applicable even when T is small.

3.2.4 Serial Correlation Model

In an error component model, if error terms from different (usually adjacent) time periods (or cross-section observations) are correlated, the error terms are serially correlated. Under the serial correlation assumption, u_{it} is correlated with u_{is} 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 an restrictive assumption for economics relationship, such as the ones in investment and production demand forecasting. In serial correlation models, the error terms 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). Observe that 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. (Errors might also be lagged, e.g. if the data points are collected quarterly, the errors in summer of a particular year might be correlated with the errors of summer in the next year.) With positive serial correlation, errors in one time period would be positively correlated with errors in the next time period.

If the serial correlation is present, the error term ϕ_{it} can be expressed as follows (Baltagi 2008)

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

where $|\rho|<1$ and $E(\mathrm{u}_{it})=0\,, E(\mathrm{u}_{it}\mathrm{u}_{it}^{'})=\sigma^2$. 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 $|\rho_2|<1$, $|\rho_1|<(1\text{-}\rho_2)$ and $E(\mathrm{u}_{it})=0\,, E(\mathrm{u}_{it}\mathrm{u}_{it}^{'})=\sigma^2$. When $\rho_2=0$, this model follows an AR(1) process. Arellano and Bond (1991) and Baltagi and Li (1992) considered this serially correlated structure in the error components model. If the one-way error component model following 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 $E(\mathrm{u}_{it})=0\,, E(\mathrm{u}_{it}\mathrm{u}_{it}^{'})=\sigma^2$.

This can be extended to the MA (q) case and the autoregressive moving average ARMA (p, q) case on ϕ_{it}. Drukker (2003) gave a detailed illustration on how one could test serial correlation scientifically. Notice that 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. Serial correlation will not affect the unbiasedness or consistency of the OLS estimators, although it does affect their efficiency (Baltagi 2008). The first-differenced GMM estimator for the AR (1) panel data model was investigated by Holtz-Eakin et al. (1988) and Baltagi and Li (1991). Besides, GLS estimator was also adopted in estimating serial correlation panel data models in (Frees and Miller 2004). To be specific, Frees and Miller (2004) used serial correlation panel data model (so called longitudinal data models) to predict the sales of state lottery tickets. Using the mean absolute error criteria and the mean absolute percentage error criteria, the best forecasts were given by the error component model with AR(1) disturbances followed by the fixed-effects model with AR(1) disturbances. Baltagi et al. (2007) considered 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 revealed that ignoring these correlations might result in misleading inference. Serial correlation that exists among the data sets that are collected repeatedly across time occurs in time-series studies when the errors associated with a given time period would carry over into future time periods.

3.2.5 Summary

This section introduced the common used panel data models for sales forecasting and summarized the corresponding estimation methods for each category. The OLS estimator is unbiased and consistent for static model with both fixed-effects and random-effects models. However, the situation is different if the model contains exogenous variables in addition to the lagged dependent variables. The OLS estimator is no longer efficient for dynamic cases, while MLE and GMM estimators are suggested to be useful for both fixed-effects and random-effects dynamic models estimation in the literature. The proper estimators for different panel data models are summarized in Table 3.5.

Table 3.5 Efficient estimators for different panel data regression models

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

3.3 Tests

In the above sections, several panel data regression models which were commonly used in forecasting problems and the corresponding estimation methods were introduced. However, how to decide the proper regression model under different industry settings and how to decide the individual effects and cross-section or time-series dependence require further explorations. In fact, a series of pretests are important to address these issues. In this section, the testing methods that can help to choose the suitable panel data models would be discussed.

3.3.1 Panel Stationary Tests

For panel data applications it is important to know whether an observed panel series is stationary or nonstationary. In the past decades, several testing approaches were proposed; one of them was the "panel unit root test". The earlier literature on panel unit root test all assumed that the individual time series in the panel were cross-sectionally independent. For example, Quah (1994) proposed the asymptotically normal tests for a unit root. Levin et al. (2002) devised an adjusted t-test (LLC) for a unit root for various panel data models. Im et al. (2003) proposed the Fisher-ADF and Fisher-PP tests for examining the unit roots in heterogeneous panels. Baltagi and Kao (2001) gave 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 tests that account for the cross-sectional correlation were proposed in the literature (Bai and Ng 2004; Breitung and Das 2005; Breitung and Das 2008; Chang 2002; Choi and Chue 2007; Gengenbach, et al. 2009; Phillips and Sul 2003). In particular, Bai and Ng (2004) studied whether the difference in finite-sample properties could be used to trace how the pooled autoregressive coefficient is estimated.

3.3.2 Individual-specific Effects Test

Testing for the correlation of unobservable individual effects with the right-hand-side variables in panel data regressions is a widespread practice (Arellano 1993). Considering a common panel data model as described in Eq. (2), the individual-specific effects (α_i) among cross-section individuals are unobserved and may be correlated with X_{it}. Generally, these effects are treated as fixed effects or random effects. For the fixed effects model, 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 the random-effect model treats the effects as random variables. When deciding between these two effects in a panel data model, a Hausman (Hausman 1978) pretest, with the assumption that the random effects 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 help to make the choice between the random-effects and fixed-effects estimators (Baltagi, et al. 2003). Hausman (1978) proposed an asymptotic chi-square 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 there is a reason to doubt the assumption in which the efficient estimator is based. This null hypothesis frequently does not withstand empirical scrutiny since a situation usually occurs that should be considered “an exception rather than the rule” (Frondel and Vance 2010). Then, Hausman and Taylor (1981) proposed a model that introduced an instrumental variable estimator using both between- 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) suggested 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) and Breusch and Pagan (1980), could also be employed to test the individual effects for panel data model. As a remark, the extensive Monte Carlo on testing in this error component model was performed by (Baltagi, et al. 1992).

3.3.3 Spatial Correlation Test

As described earlier, the standard panel data model assumes that no spatial correlation exists. However, there are real world cases in which this might not be true. For example, for trade flows across a panel of countries, there might be spatial effects affecting this trade, depending on the distance between these countries. The panel data model should hence be developed by considering spatial correlation. In spatial dependence models, the structure of economic distance measurement provided by cross-sectional data is similar to that provided by the time index in time series. Several estimating methods for spatial correlation models, such as maximum likelihood methods, and generalized method of moments (GMM), were examined in the above section. As mentioned before, 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 would be reviewed. Anselin (1988, 2001), Anselin and Bera (1998) and Kelejian and Robinson (1998) developed LM (Lagrange Multiplier) tests for spatial autocorrelation in cross-sectional spatial data which are observed for a given time point. Then, Baltagi et al. (2003) extended the Breusch and Pagan LM test to the spatial error component model and derived a conditional LM test to testing for random region effects in the panel as well as spatial correlation across these regions, which is able to test for random regional effects given the presence of spatial error correlation and also spatial error correlation given the presence of random regional effects. Recently, Jang and Shin (2014) suggested the joint LM and LR tests, the marginal LM and LR tests, and the conditional LM and LR tests, for testing both spatial correlation and time effect. Their limiting null distributions were also derived by conducting a Monte-Carlo experiment. Following Baltagi et al. (2003)'s work, Gengenbach et al. (2009) derived a joint LM test which could simultaneously test for the absence of spatial lag dependence and random individual effects in a panel data regression model. This joint LM test allows for spatial lag dependence of the autoregressive kind in the dependent variable rather than the error term.

3.3.4 Serial Correlation Test

Testing for serial correlation has been a standard practice in applied econometric analysis because 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, the distant literature (Breusch 1978; Breusch and Pagan 1980; Godfrey 1978; Banerjee, et al. 1998) already studied serial correlation testing problems widely in the past decades. For panel data, Gardner (1960) first extended the error component model to take into account serial correlation in the remainder disturbance term and test for serial correlation, assuming there are no random effects. Then, Bhargava et al. (1982) modified the Durbin-Watson statistics (Bhargava, et al. 1982) to test for serial correlation when the individual effects are assumed to be fixed. Baltagi and Li (1991) derived a simple (LM) test which jointly tests the presence of random individual effects and serial correlation. Baltagi and Li (1995) also addressed this kind of joint testing problem. By generalizing the testing methods for time series data, Li and Hsiao (1998) proposed two methods to test zero first-order serial correlation, higher-order serial correlations in a residual-based dynamic panel data model. Some Monte Carlo experiments were conducted to examine the finite sample performances of the proposed tests. After that, Wooldridge (2002) developed a new test for serial correlation in random-effects or fixed-effects one-way models. It is a very useful test because it can be applied under general conditions and is easy to implement. Baltagi et al. (2007) generalized the previous studies by deriving test statistics for the spatial panel data model with serial correlation. Extending the time series test in Baltagi et al. (2007), Westerlund (2007) proposed a serial correlation test for panel data based on the structural properties that do not impose any common factor restriction. The simulation results suggested that the proposed test had good small-sample properties with small size distortions and high power relative to other popular residual-based panel cointegration tests.

3.4. Panel Data Model Selection

As discussed above, since there are many different kinds of analytical panel data models and estimators available, how to select the right model and testing method becomes a critically important issue. Based on the analysis conducted in the above sections, a novel flowchart based process for decision makers to identify the right panel data forecasting model and estimator for the analysis is shown in Figure. A3.1. Details of the flowchart are shown in Figure A3.2. This particular process is important in helping decision makers identify the right panel data based models for conducting fashion sales forecasting with respect to the data formats and company requirements.

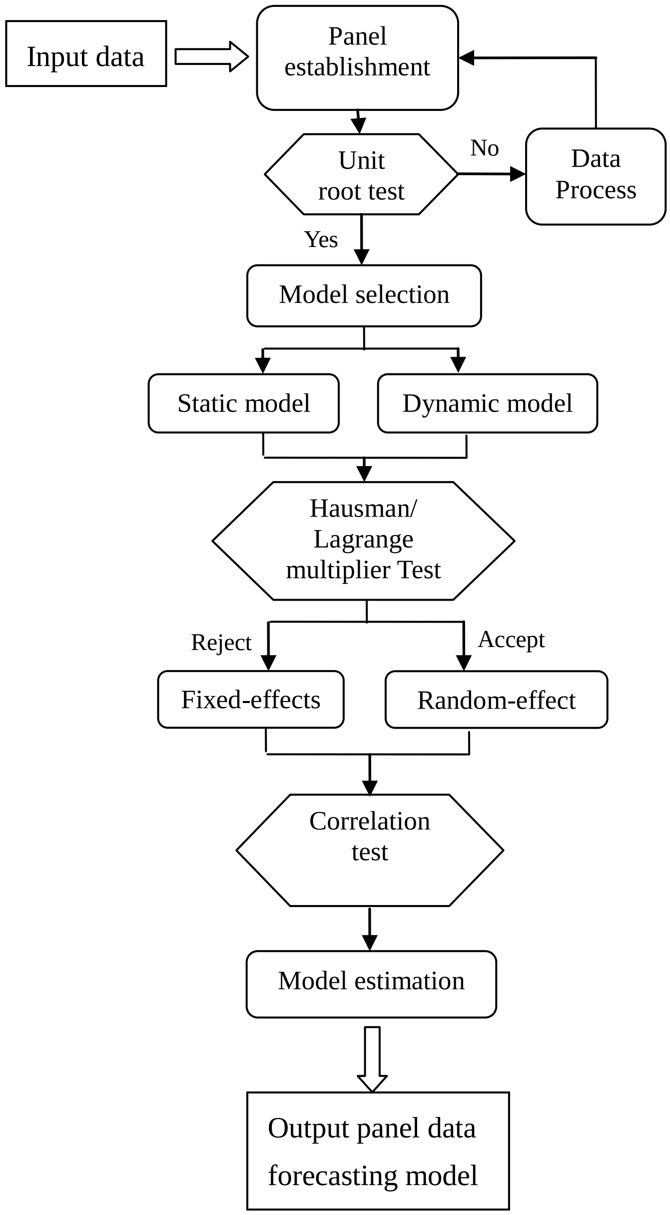


Figure A3.1 A novel flowchart developed in this paper for establishing the right panel data model for industrial applications.

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) (Banerjee 1999) (Kao and Chiang, 2001). For applications, please refer to (Lee 2005) (Costantini and Martini, 2010) (Li et al., 2013) (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/Lagrange multiplier 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 (2008).

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.

Figure A3.2 The novel panel data model selection process.

Chapter 4 A Comparative Study for Fashion Sales Forecasting Models[4]

Many kinds of fashion sales forecasting models were examined in Chapter 2, and the versatile panel data models were also reviewed extensively in Chapter 3. This chapter proceeds to report a comparative study among these models. To be specific, the panel data model and different kinds of commonly seen models for fashion sales forecasting are introduced. Then, using a dataset of real world sales data in fashion retailing, a comparison among these models is made. Finally, results of an AHP analysis, based on an industrial survey, are reported to show the preference of fashion industrialists regarding these fashion sales forecasting models.

As a remark, in this chapter, the datasets used to conduct this comparative study are weekly real sales data from a fashion boutique in Hong Kong. The time period of this datasets covers 9 months and in total six fashion items with seven kinds of color, together with other related properties of the items are included. Table 4.1 illustrates an extract of the original dataset. In order to examine the forecasting performance of different kinds of methods in the same scenario, the first 24 samples (about 6 months) are used as the training data (for estimating the model parameters) and the remaining 12 samples (about 3 months) are used to conduct the forecasting test.

[4] A part of this chapter is extracted to develop a paper under journal review.

Table 4.1 Sales data (an extract)

Date (year-month-day)	Item Code	Color	Quantity	Price
2005-04-16	1	black	1	89
2005-04-16	1	yellow	1	39
2005-04-16	2	brown	1	63
2005-04-16	1	red	1	76
2005-04-16	5	blue	1	79
2005-04-16	1	black	1	69

4.1 The Panel Data Forecasting Model

As reviewed in Chapter 3, the panel data combines time-series and cross-sectional data sets. Either time series or cross-sectional data is a special case of panel data in one-dimension only. Mathematically, the panel data has *I* cross-sections (i.e. number of fashionable item) and *T* time intervals. The common panel data forecasting model is represented as：

$$y_{it} = \alpha + X_{it}^{'}\beta + u_{it} \quad i = 1,...,N; t = 1,...,T \tag{17}$$

with i denoting individuals, the fashion items in our case, and t denoting time; the subscript i, therefore, denotes the cross-section dimension whereas t denotes the time-series dimension. α is a scalar, β is $K \times 1$ and X_{it} is the it th observation on K explanatory variables. u_{it} is the error item.

In panel data forecasting models, it is impossible to put all influence factors into the model and only some most important factors are selected as the inputs (to make the panel data model). Since there are evidences that price is the most critical element affecting demand (Hsiao 2003; Jain and Rao 1990; Lohse, et al. 2000; Schultz 1935), it was chosen as the key influence factor to construct the panel data forecasting model. In the panel data based forecasting models, the time-series trend of previous sales; the prices of the items under study, and the whole panel impact from the other correlated products are chosen to be the decision variables for conducting sales prediction. They allow us to assess the effects of price, previous sale and correlated items on the sales of fashion product. Thus, from a theoretical point of view, the sales for each fashion product will be a function of the sales quantity of the last period, the corresponding price, the influence patterns from the other correlated items and the nonlinear uncertainty during the same period. The dynamic panel data forecasting model is constructed as：

$$S_{it} = \alpha_i^* + \gamma S_{it-1} + \beta \cdot \mathrm{P}_{it} + \mu_{it}, \quad i = 1,...,I; t = 1,...,T, \tag{18}$$

where S_{it} is the sale of item i during the time interval t; P_{it} is the corresponding price; β is coefficient for cross-section; γ is coefficient for time-series; The independent error term μ_{it} distributes over i and t, with mean zero and variance σ_u^2 and is assumed to be uncorrelated with price and previous sale. Notice that Eq. (18) states the relationship among different decision variables, i.e. how the sales are related to the previous sale and its corresponding price. The forecasting result of the panel data based method can be obtained by using the maximum likelihood estimator. For the lag selection, in this paper, the simple case with "one lag" was employed as the goal of this thesis was to establish an implementable panel data based application. The optimal choice of "lag" will be postponed to future research.

The advantage of having panel data as compared to cross-section sets is that it allows us to test and relax the assumptions that are implicit in cross-sectional analysis (Hsiao 2003). As mentioned in Chapter 3, before forecasting the sales of fashionable products by using panel data model, a series of tests should be conducted based on real sales data firstly. The modeling process of panel data forecasting model is illustrated in the following chart (Figure. 4.1).

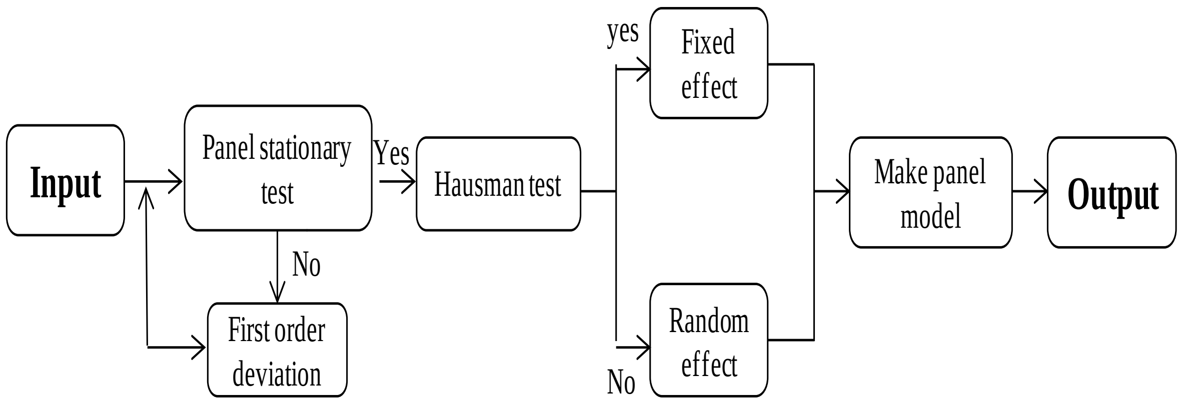


Figure 4.1 The panel data modeling process.

As described in Chapter 3, to establish a three-dimension relationship model among sales, previous sales and corresponding price, it is necessary to test whether the panel data is stationary or not. Table 4.4 provides the testing results which imply that the probability of having a common unit root is 0. The null hypothesis that the common unit root of sale and price series is non-stationary is hence rejected. In other words, the panel datasets that we adopt to construct forecasting model are stationary and the panel data estimation model can be derived directly.

Table 4.4 Unit root test result

Method	Null hypothesis	Statistic	Probability
Levin,Lin&Chu t*	common unit root process	-12.0064	0.0000
ADF- Ficher Chi-square	individual unit root process	178.015	0.0000
PP- Ficher Chi-square	individual unit root process	199.606	0.0000

*Significant at the 5% level.

Table 4.5 Hausman test result

Test Summary	Chi-Sq. Statistic	Chi-Sq.d.f	Probability
Cross-section random	128.4	2	0.0000

*Estimated cross-section random effects variance is 0.

Following the unit root test, Hausman test is conducted to examine the individual effects, fixed-effects or random-effects. The Hausman testing result further reveals that a fixed effect model should be constructed for the panel-data forecasting problem.

Thus, the linear component of fashion product sales can be described as follows:

$$S_{it} = m + \alpha_i^* + \gamma S_{it-1} + \beta \cdot \mathrm{P}_{it} + \mu_{it}, \quad i = 1,\ldots,N, t = 1,\ldots,T, \tag{19}$$

where m represents the effect from the whole panel, α_i^* represents the effects of those variables peculiar to the i th item in more or less the same way over time., the parameters γ and β indicate the degree that sale of item i at time t is determined by the value of the previous sale and the corresponding price.

Notice that Eq. (19) states the relationship among different variables, i.e. how the amount of product sales is related to the previous sales and the corresponding price. The estimation of Eq. (19) was done by using the maximum likelihood estimator. Table 4.6 summarizes the outcome of the estimation procedure using a panel of data.

Table 4.6 Estimation result of the sale forecasting model

coefficient	m	γ	β	α_1	α_2	α_3	α_4	α_5	α_6
estimation	5.49	0.15	0.02	15.4	-5	-4.3	3.01	-4.8	-4.3
T-Statistic	5.44	1.70	1.41	-	-	-	-	-	-

The estimation results show that the sales of each fashion item mainly depends on the overall sales trend m. Previous sales and the corresponding price are also important for explaining the sales changing of each item. This result is consistent with our expectation because the price of fashion product should play an important role.

4.2 Other Commonly Used Fashion Sales Forecasting Models

4.2.1 Statistical Models

In conducting sales forecasting for fashion products, the classical approach starts by analyzing the product features (Nenni, et al. 2013). After that, fashion companies need to determine the forecasting approach. Being quick, intuitive and easy to apply, statistical methods such as Auto Regression Integrated Moving Average (ARIMA) and Seasonal Auto Regression Integrated Moving Average (SARIMA) approaches are commonly used for quick sales forecasting. For the use of these classical simple models, please refer to other papers, such as (George 1994; Liu, et al. 2013; Thomassey 2014). In this chapter, the ARIMA model is employed for the comparison study.

4.2.2 Extreme Learning Machine

For fashion sales forecasting, another popular set of forecasting models for predicting fashion product's demand is by artificial neural networks (ANN) (Hamzaçebi, et al. 2009). It is well known that even a simple ANN would take a substantial amount of time to complete a forecasting task (e.g., it may take several minutes, and evolutionary neural networks (ENN) may take hours (Au, et al. 2008)). The long computational time becomes a major barrier for the deployment of many ANN and ENN based forecasting models in real world fast fashion sales forecasting. Relatively recently, there is an invention of a fast single-hidden layer feed-forward neural network (SLFN), called the extreme learning machine (ELM) (Pao, et al. 1994; Hsu and Wang 2007; Huang, et al. 2006; Rong, et al. 2008; Sun, et al. 2008; Sun, et al. 2007; Xia, et al. 2012). ELM is able to learn much faster than many conventional gradient-based learning methods reported in the classical neural networks literature. As a remark, Sun et al. (2008) is the first piece of work which applied ELM in conducting fashion sales forecasting. After that, a lot of studies emerged. In the following, this pioneering sales forecasting analytical model would be introduced. For more details (including detailed illustrations and Figures), please refer to (Sun et al. 2008).

As mentioned above, the ELM is a SLFN with the inputs of variables x_{ij}. By its nature, it randomly assigns the input weight matrix $\mathbf{W}$, and analytically determines the output weight matrix $\boldsymbol{\beta}$. To be specific, suppose that one wants to train the "SLFN" of the ELM with K hidden neurons and an activation function vector $\mathbf{g}(x) = (g_1(x), g_2(x), \cdots, g_K(x))$ to learn from N distinct samples (x_i, t_i), where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R_n$ and $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in R_m$. If the SLFN in the ELM can approximate these N samples with a zero error, then we have $\sum_{j=1}^{N} \| \mathbf{y}_j - \mathbf{t}_j \| = 0$, where $\mathbf{y}$ is the SLFN's output.

In the ELM, the parameters $\boldsymbol{\beta}_i$, $\mathbf{w}_i$ and $\mathbf{b}_i$ satisfy the following system of equations:

$$\sum_{i=1}^{K} \boldsymbol{\beta}_i g_i(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{t}_j \text{ , } j = 1, \cdots, N \text{ ,} \tag{20}$$

where

$\boldsymbol{\beta}_i = [\beta_{i1}, \cdots, \beta_{im}]^T, i = 1, \cdots, K$ links the ith hidden neuron and the output neurons, $\mathbf{w}_i = [w_{i1}, \cdots, w_{in}]^T$ is the weight vector linking the ith hidden neuron and the input neurons, and b_i is the ith hidden neuron's threshold.

Note that in the ELM, by default, the input weights and hidden biases are all randomly generated instead of tuned. As a result, one can determine the output weights by finding the least-square solution to the given linear system of equations.

To employ ELM in fashion sales forecasting, from Sun et al. (Sun et al. 2008), the following steps apply:

Step 1) Get the sales data, and select the factors that have significant effects on the product demand as the inputs of ELM; note that the statistical analysis conducted in Ren et al. (2014) can be used to identify the factors which have significant effects on demand.

Step 2) With the given dataset, divide the data into training data, testing data, and forecasting data sets randomly. Normalize the training data and the testing data, and select the activation function of hidden neuron and choose the neuron number of hidden layer of ELM;

Step 3) Input training data and testing data, compute the outputs of ELM, un-normalize the outputs, then obtain the predicted sales time series of the training data and the testing data;

Step 4) Based on the input and output weights obtained by Steps 2 and 3 above, compute the predicted sales time series and the corresponding predicting error.

Note that even though ELM runs much faster than the classical ANNs and ENNs, it still requires a certain amount of time to complete the sales forecasting task and it also requires a sufficient amount of data for training in order to yield good sales forecasting results.

4.2.3 Grey Model

To conduct time-series sales forecasting with insufficient historical data, the grey model (GM) has been known to be a very good candidate (Chen and Ou 2009; Choi, et al. 2012; Hsu and Wang 2007; Lei and Feng 2012; Li and Xie 2014; Lin and Lee 2007; Wang 2014; Xia and Wong 2014). Observe that GM was derived from the systems science literature which proposes that a system often faces uncertainty, and it is often difficult, if not impossible, to classify the system purely as “black” or “white”. Thus, based on this argument, Deng (1989) cleverly defined a system which has both “known” and “unknown” information as a grey system.

In the analytical model, a GM is usually represented by GM (l, k), where l is the order of differential equations employed, and k is the number of variables in the GM. Note that the simplest yet the most commonly used GM is GM (1,1), which is called the single-variable first-order grey model (SFGM) (Li and Xie 2014; Li, et al. 2011; Nenni, et al. 2013). In (Choi, et al. 2012), the SFGM has been used to conduct time-series forecasting for fashion demand and their analytical model is described as follows.

First, in the time series analysis by using SFGM, the original demand time series is represented by $X^0 = [x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(\mathrm{n})]$, where $x^{(0)}(i)$ is the demand data point of time series at time i (i=1,2,…,n). With X^0, one can get a new “aggregated demand time series” X^1 by the following simple operation:

$$X^1 = [x^{(1)}(1), x^{(1)}(2), ..., x^{(1)}(\mathrm{n})], \text{ and } x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i). \tag{21}$$

Under the SFGM based demand forecasting method, the forecasting of X^1 at time k, given by $\hat{x}^{(1)}(k)$, can be derived using the method in Deng (1989), and the forecasted future demand value of X^0 at time k + 1 can be found by the following:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k). \tag{22}$$

Undoubtedly, the SFGM is a simple and easy to apply forecasting method. It, together with other GM based methods, is suitable for conducting time series sales forecasting with very few historical data (which is commonly the case in fast fashion business operations). Thus, GM is a good candidate with which one can develop fashion sales forecasting applications. However, kindly note that the SFGM and other GM based methods are known to be unreliable. This is especially true for those time series which are highly volatile.

4.3 Comparisons of Different Fashion Forecasting Models

After the introduction of the fashion sales forecasting models above, their performances are examined in terms of the following five evaluating criteria: Accuracy, speed, data sufficiency requirement, stability, and ease of implementation. To differentiate from the hybrid panel data based model that will be examined in Chapter 5, the panel data based model in this chapter, is called the pure panel data model (PPD).

A. Accuracy

First of all, the most critical measure to compare is accuracy. In order to have a fair and scientific comparison, the real datasets presented in Section 4.1 are employed.

In this comparison analysis, Mean squared error (MSE) and symmetric mean absolute percentage error (SMAPE) are used to measure the forecasting accuracy of our proposed model. Notice that the MSE is a popular way to quantify the difference between values implied by an estimator and the true values of the quantity being estimated (Lehmann and Casella 1998). The MSE measures the average of the squares of the "errors." The error is the amount in which the value implied by the estimator deviates from the quantity to be estimated. It can be described as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(A_t - F_t)^2 , \tag{23}$$

where F_t is a vector of *n* predictions, and A_t is the vector of the true values.

To supplement MSE, in this comparison study on accuracy, the SMAPE is also employed. Notice that the SMAPE measures the relative error rather than the mean absolute percentage error (MAPE) and it is usually defined as follows (Armstrong and Forecasting 1985):

$$SMAPE = \frac{\sum_{t=1}^{n} | F_t - A_t |}{\sum_{t=1}^{n} (F_t + A_t)} . \tag{24}$$

Further observe that SMAPE allows measuring the direction of the bias in the data by generating a positive and a negative error.

In this study, ARIMA, ELM and GM were chosen as comparison models with the studied PPD model because they were well-examined in the literature and exhibited the following features: 1. ARIMA is the most commonly used, easy to implement, one-dimensional time-series statistical method for sales forecasting; 2. ELM is very fast in computation and it is able to learn and operate much faster than many conventional gradient-based learning methods reported in the classical neural networks literature. 3. GM is a very good method to conduct sales forecasting with insufficient historical data. The forecasting results of the four models under examination, ARIMA, GM, ELM and PPD, are illustrated in Table 4.7. It is obviously that the panel data model outperforms all the other three models for almost all studied items, with the exception for item 1 in which it does not perform as well as ARIMA.

Table 4.7 Forecasting performances in accuracy of ARIMA, GM, ELM and PPD

Item	**ARIMA**		**GM**		**ELM**		**PPD**	
	MSE	SMAPE	MSE	SMAPE	MSE	SMAPE	MSE	SMAPE
1	62.8	9.8%	154.2	20.2%	86.9	12.4%	64.5	10.6%
2	3.4	29.8%	4.0	35.6%	4.8	40.1%	3.0	27.4%
3	18.1	44.9%	1.6	23.4%	3.9	40.3%	1.0	17.4%
4	19.6	16.1%	44.5	28.1%	19.1	15.1%	13.1	10.8%
5	4.8	42.8%	5.9	55.4%	4.0	38.4%	3.3	30.3%
6	2.4	48.3%	2.8	56.7%	2.5	41.4%	1.8	34.4%
Mean	18.9	32.0%	35.5	36.6%	20.2	31.3%	14.5	21.8%

B. Speed

First of all, note that all these methods are known to be able to yield forecasting result in a timely manner and hence they are "fast". While if we go deeper in terms of how fast each method performs, we can see that in terms of speed, PPD and ARIMA are all fastest (in seconds) because they are statistical methods. ELM and GM are also fast but definitely not as fast as ARIMA and PPD.

C. Data sufficiency requirements (DSR)

In terms of data requirements, some methods have higher demand for data sufficiency than the others. To be specific, ELM needs to have sufficient data in order to do a good job. Interestingly, PPD, ARIMA, and GM have much less demand for having a lot of data because: (i) For PPD, the panel-data puts high emphasis on correlation related information. Thus, there is no need to have a lot of historical data for each item in making a sound forecasting; (ii) For ARIMA, it is simplest and can be applied even if the number of data points is very little. (iii) For GM, it is known to be functional even in the absence of enough data.

D. Stability

For stability of forecasting results, the case is rather clear. First of all, PPD and ARIMA are known to generate stable and reliable forecasting results as they are purely statistical methods. However, ELM and GM are known for their shortcomings in terms of yielding unstable forecasting results. Thus, PPD and ARIMA based demand forecasting models have high stability whereas ELM and GM have low stability.

D. Ease to use

For the issue on whether the forecasting model is easy to use and implement, we find that ARIMA requires only simple analytical closed-form relationship to conduct forecasting and PPD requires only the basic regression of the available panel data. Both of them can be done automatically by many commercial software packages. Thus, both ARIMA and PPD are easiest to implement and use in practice. For ELM and GM, they all can conduct demand forecasting in an automatic way after implementing the respective algorithms. Thus, they are also easy to use from that sense but should still be less easy to use compared to ARIMA and PPD.

Table 4.8 Comparisons among the reviewed models

Methods	**Five factors**				
	Speed	**DSR**	**Stability**	**Use**	**Accuracy**
ARIMA	Fastest	Low	High	Easiest, intuitive	Medium
PPD	Fastest	Low	High	Easiest, intuitive	Highest
ELM	Fast	High	Low	Easy	Medium
GM	Fast	Low	Low	Easy	Lowest

From the above investigation, the strengths and the weaknesses of the fashion sales forecasting models are found. As a summary, Table 4.8 shows the item-to-item systematic comparison among the reviewed fashion sales forecasting models. From Table 4.8, among these important models, it is crystal clear to observe that PPD is especially versatile and helpful for developing fashion sales forecasting systems.

4.4 Industrial survey & AHP analysis

Following the above comparison analysis, an industrial survey was conducted to learn about how human decision makers evaluate these fashion forecasting methods. In this part of the study, firstly, the importance of different analytical models will be compared by considering human decision makers' evaluation. After that, how different groups of human decision makers evaluate the forecasting systems is further analyzed.

4.4.1 Analytic Hierarchy Process (AHP) Analysis

A very commonly used multiple criteria decision-making tool, called the AHP analysis (Vaidya and Kumar 2006), is used to conduct the comparison study of forecasting systems. In the AHP analysis, the factors that are much important for decision making are structured in a hierarchy structure, which is composed of goal, criteria, sub-criteria and alternatives. For more details, refer to Saaty (1980).

In this section, the four types of sales forecasting systems, ARIMA, ELM, GM and PPD, would be compared with respect to the five factors by using inputs from industrialists by conducting the AHP analysis. To be specific, the first level is called the ‘goal’ level, the second level is the ‘criteria’ level, and the third level is called the ‘alternative’ level. Figure .4.2 depicts this structure.

In the criteria level, an industry survey was conducted to obtain the relative weight for each criterion element. For this survey, the respondents were identified by a convenience sampling method with the author’s own network. All respondents had worked in or were working in the fashion industry at the time of survey. The respondents had diversified backgrounds and occupied different positions from fashion companies selling different products and serving different target markets. In total, 123 questionnaires were collected and 114 questionnaires of them were valid (see the questionnaire in Appendix A). According to the collected survey data, the weights of all forecasting measurement in criteria level were calculated as shown in Table A4.1.

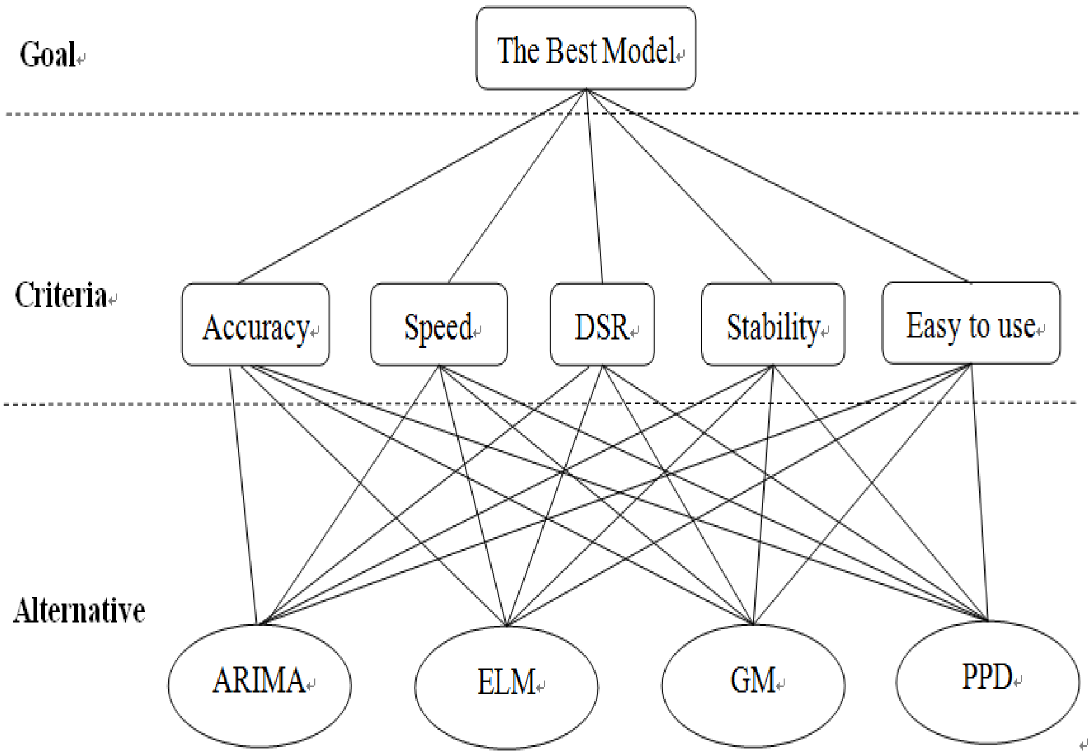


Figure 4.2 The AHP Structure.

Table A4.1 Weights of forecasting performance in criteria level

Criteria	Accuracy	Speed	DSR	Stability	Ease to use
Weight	0.214	0.188	0.195	0.208	0.196

Table A4.2 Weights of forecasting performance in different groups

Groups	**Category**	**Accuracy**	**Speed**	**DSR**	**Stability**	**Ease to use**
Position	manager	0.212	0.189	0.196	0.204	0.199
	operator	0.215	0.188	0.194	0.209	0.194
Production	fashion	0.216	0.193	0.190	0.205	0.195
	basic	0.212	0.183	0.199	0.209	0.196
Market	high	0.218	0.191	0.191	0.210	0.191
	low	0.213	0.186	0.196	0.206	0.198

In order to conduct the comparison study, the questionnaires inputs were classified into different groups such as the position group (manager and operator), the product group (fashion product and basic product) and the targeting market group (high-market and low-market). The weights of forecasting measurement criteria in different sorting groups were given in Table A4.2. From the analysis, accuracy is found to be the most important criterion (among the five criteria) in terms of demand forecasting performance measures. This result is intuitive and consistent with the literature.

In the alternative level, the comparison matrices of each forecasting model measured by different criteria are designed by expert choice (Tables A4.3-4.7) and the corresponding weights are calculated as shown in Table A4.8 following the standard computational process.

Table A4.3 Accuracy

Models	ARIMA	GM	ELM	Panel data
ARIMA	1	3/1	3/5	3/5
GM	1/3	1	1/5	1/5
ELM	5/3	5/1	1	5/5
Panel data	5/3	5/1	5/5	1

Table A4.4 Speed

Models	ARIMA	GM	ELM	Panel data
ARIMA	1	5/3	5/3	5/5
GM	3/5	1	3/3	3/5
ELM	3/5	3/3	1	3/5

Panel data	5/5	5/3	5/3	1

Table A4.5 DSR

Models	ARIMA	GM	ELM	Panel data
ARIMA	1	3/1	3/5	3/3
GM	1/3	1	1/5	1/3
ELM	5/3	5/1	1	5/3
Panel data	3/3	3/1	3/5	1

Table A4.6 Stability

Models	ARIMA	GM	ELM	Panel data
ARIMA	1	5/3	5/1	5/5
GM	3/5	1	3/1	3/5
ELM	1/5	1/3	1	1/5
Panel data	5/5	5/3	5/1	1

Table A4.7 Ease to use

Models	ARIMA	GM	ELM	Panel data
ARIMA	1	5/1	5/1	5/5
GM	1/5	1	1/1	1/5
ELM	1/5	1/1	1	1/5
Panel data	5/5	5/1	5/1	1

Table 4.8 Weights of the four models

Criteria	ARIMA	GM	ELM	Panel data
Accuracy	0.214	0.072	0.357	0.357
Stability	0.3125	0.1875	0.1875	0.3125
DSR	0.25	0.083	0.417	0.25
Speed	0.357	0.214	0.072	0.357
Easy to use	0.417	0.083	0.083	0.417

Combining weights for the criteria level in Table A4.1 and weights for the alternative level, the relative contribution of each forecasting model with respect to the evaluating performance is shown in Table A4.9.

Table A4.9 The relative contribution of each forecasting model

Category	ARIMA	GM	ELM	PPD
Overall	0.3090	0.118	0.224	0.3400
Manager	0.3092	0.1176	0.2241	0.3338
Operator	0.3088	0.118	0.2241	0.3433
Fashion	0.3085	0.1178	0.2231	0.3394
Basic	0.3087	0.1176	0.224	0.3390
High-market	0.3087	0.1184	0.2239	0.3390

From Table A4.9, it is easily to observe that PPD shows the best performance with the consideration of all five evaluating criteria (including forecasting accuracy, speed, data sufficiency requirements (DSR), stability, and ease of usage), while GM performs the worst. In summary, considering the comprehensive performance and decision makers' preference, PPD is the most versatile and helpful model for conducting demand forecasting from the perspective of the practitioners.

4.4.2. Further Analysis

A. Correlations of forecasting criteria

The correlation testing results (listed in Table A4.10) indicate that when decision makers rank the importance of each forecasting criteria, there are significant correlations among different measurements except for "accuracy" and "ease to use". In other words, the grading score of 'accuracy' is not significantly affected by "ease to use".

Table A4.10 Correlation testing results

Criteria	Accuracy	Speed	DSR	Stability	Ease to use
Accuracy	1	0.360**	0.415**	0.407**	0.167
Speed	0.360**	1	0.440**	0.495**	0.557**
DSR	0.415**	0.440**	1	0.481**	0.304**
Stability	0.407**	0.495**	0.481**	1	0.508**
Ease to use	0.167	0.557**	0.304**	0.508**	1

**. Correlation is significant at the 0.01 level (2-tailed).

B. Comparisons analysis

In order to learn more about how human decision makers evaluate forecasting methods, the mean importance of each forecasting criterion graded by them have been compared in different sorting groups (including position group, product group and targeting market group). The following three Figures illustrate the mean importance of each criterion.

From Figure A4.2, it is easy to find that for both the manager (senior) and the operator (junior) groups, the criterion "accuracy" is viewed as the most critical one among all the five criteria. Interestingly, speed is treated as the least important one. From Figure A4.3, for practitioners working in the fashion companies which mainly sell highly fashionable items, DSR is treated as the least important criterion whereas speed is treated as the least important criterion for the practitioners working in the fashion companies selling basic items. This difference can be explained by the fact that for the fashion companies selling highly fashionable items, they naturally have to conduct forecasting in the absence of enough data as their products are highly fashionable and short-lived. Similar to the cases reported above, “accuracy” is still the most important criterion for both the “fashion” and “basic” groups of companies. It is interesting to note from Figure. A4.4 that the practitioners working in the high market fashion companies treat DSR, together with speed, as the least important factor, whereas the practitioners working in the low market fashion companies treat speed as the least important factor. Overall speaking, one could make the observations that “accuracy” is the top most important criterion for demand forecasting from the practitioners’ perspectives whereas “speed” and “DSR” are relatively unimportant.

As a remark, since from Figure A4.2, no difference between the manager and the operator groups of practitioners could be found; thus, further analysis was conducted by dividing the group with respect to the product natures. Figure A4.5 and Figure A4.6 show the results.

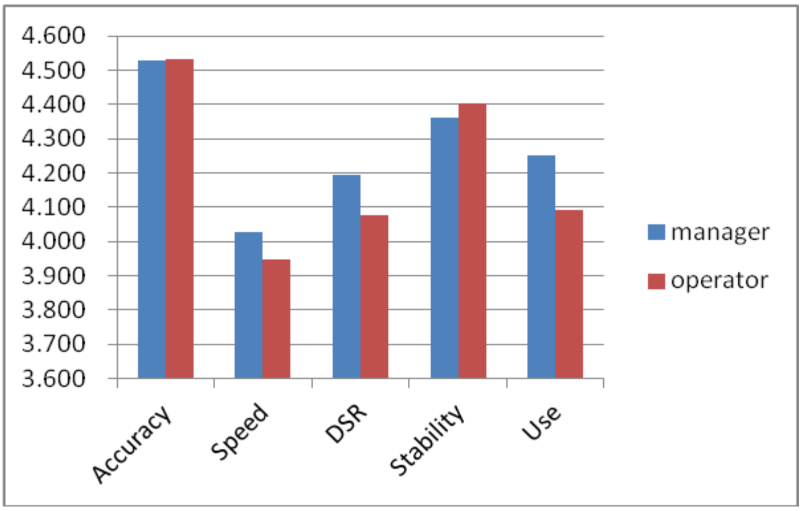


Figure A4.2 Mean importance of each criterion in position group.

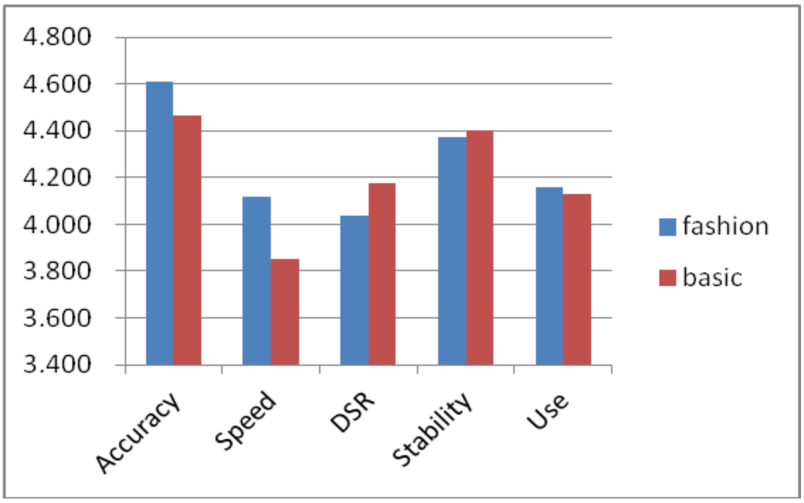


Figure A4.3 Mean importance of each criterion in product group.

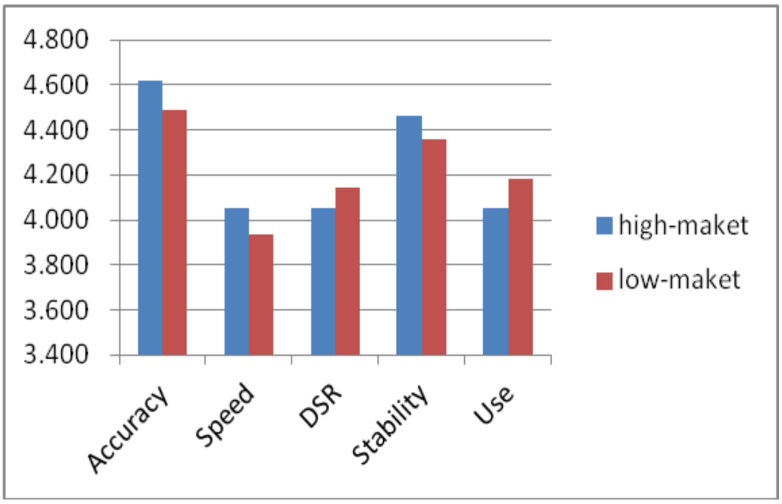


Figure A4.4 Mean importance of different criteria in market group.

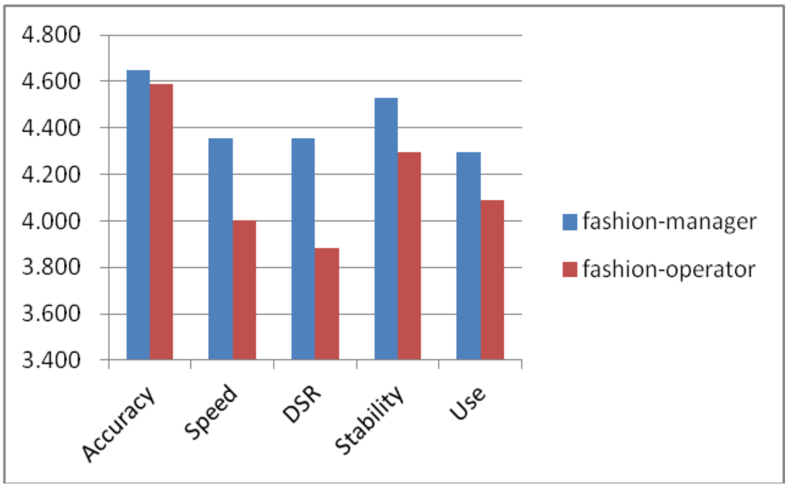


Figure A4.5 Mean importance of each criterion given by decision maker from fashion company.

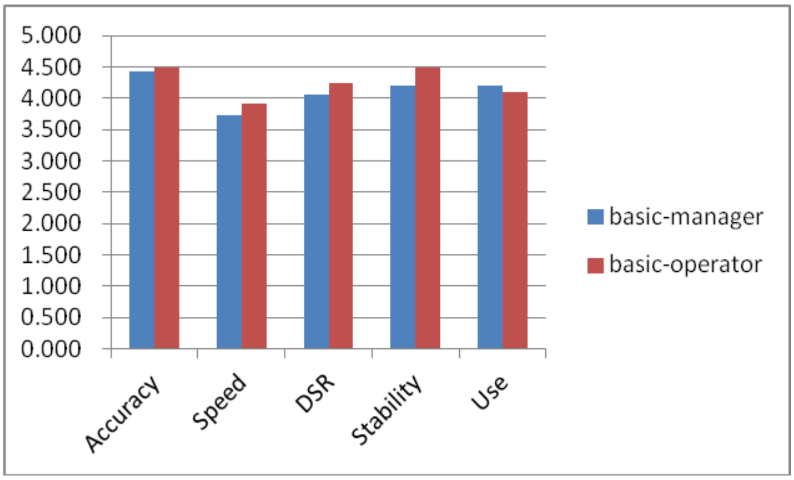


Figure A4.6 Mean importance of each criterion given by decision maker from basic company.

From Figure A4.6, it is crystal clear to observe that the mean importance scores towards each criterion for both the manager and the operator groups in the fashion companies selling basic products are very close. However, the case for the companies selling the highly fashionable products is totally different in which the mean importance scores of each criterion appears to be more different between the manager and the operator groups. To be specific, for the companies selling highly fashion items, the operators think forecasting speed and data sufficiency requirements are not as important as the other criteria, while the managers concern more about these two criteria than operators when they evaluate the performance of demand forecasting systems.

4.5 Concluding remarks

In this chapter, the strengths and the weaknesses of the four reviewed fashion sales forecasting models were discussed in five dimensions, namely accuracy, speed, data sufficiency requirements, stability, and ease to use. After that, an industrial survey was conducted to examine the industrialists' preferences towards these fashion sales forecasting systems. Based on an AHP analysis, overall speaking, it was found that accuracy would be the most important criterion among all the five studied criteria, and PPD appeared to be the champion fashion sales forecasting model from the practitioners' perspectives. Some further insights on different groups of decision makers' preferences were reported in Section 4.4.

Chapter 5 Fashion Sales Forecasting with a Panel Data-Based Particle-Filter Model[5]

The common characteristics of fashion forecasting are: a) fashion product with short life cycle and b) quick response for highly fashion trend. Fashion companies have to conduct quick prediction of the highly volatile demand of a lot of stock keeping units (SKUs). As discussed in Chapter 4, the panel data method is applicable to conduct fashion sales forecasting because of its ability in capturing the individual effects that exist among cross-sectional sales data but are not captured by the included explanatory variables (e.g., the effect from the sales of other correlated products). The advantage of such a pooling approach of forecasting has been widely demonstrated in a series of literature discussed in Chapter 3. In addition to pooling, accounting for interaction among cross-sections may prove beneficial for the purposes of forecasting. It is not difficult for us to construct a multi-dimensional relationship forecasting model using the panel data structure. The comparison analysis based on real data and industrial survey also further indicated that the panel data model is useful and versatile for conducting fashion sales forecasting. However, there is no perfect method. It is difficult for the panel data method alone to capture the nonlinear features among the variables (Hsiao 2003).

[5] A part of this chapter was published in a journal paper: Ren, S., T.-M. Choi, and N. Liu. Fashion Sales Forecasting With a Panel Data-Based Particle-Filter Model. IEEE Transactions on Systems, Man, and Cybernetics: Systems 45(3), 411-421.

In fact, the literature review in Chapter 2 indicated that, combining advantages of statistical methods and the AI methods, hybrid models were usually considered to be more efficient than the pure statistical models and pure AI models. Following this research line, a novel *panel data based particle-filter* (PDPF) is proposed in this chapter to conduct fashion sales forecasting. In the PDPF model, the time-series trend of previous sales, the prices of the items under study, and the whole panel impact from the other correlated products are chosen to be the variables for conducting sales prediction. The particle filter (PF) method is used to deal the nonlinear features that the panel data model has missed. The forecasting performance of the PDPF is further evaluated by using real data collected from the fashion industry.

5.1 Panel Data Based Particle Filter (PDPF) Model

As discussed in Chapter 1, the fashion apparel market is strongly influenced by many factors. These factors, commonly called explanatory variables, are often uncontrollable, sometimes unknown and difficult to quantify their impact (De Toni and Meneghetti 2000). It is also difficult for the panel data method alone to capture the nonlinear features among the variables. To overcome this shortcoming, we propose the use of particle filter, which is a well-known nonlinear handling method. Therefore, a hybrid model, which combines the advantages of panel data and particle filter, is proposed to investigate the complex relationship between sales amount and other influence factors in fashion sales forecasting (Sun, et al. 2008). To be specific, some key factors, such as previous sales, the corresponding price, and the interaction with the correlated product items, are all incorporated into the model. In addition, a two-component structure for the hybrid PDPF fashion sales forecasting model, namely the linear and non-linear components, is constructed. The motivation of this combination is that panel data can better describe the data characteristics from both time-section and cross-section aspects, and PF has the capability of capturing the nonlinear influence from the external environment such as the fashion trend, weather, holiday, etc. As a remark, PF is a state-space model which was known to be a powerful tool in modeling and forecasting dynamic systems (Do Chung, et al. 2012). Arulampalam et al. (2002) suggested that, in highly nonlinear environments, a nonlinear filter such as a particle filter can offer a good performance in tracking unexpected changes. In this chapter, PF is adopted to predict the uncertain patterns since PF is particularly useful in dealing with nonlinear and non-Gaussian problems (Gordon, et al. 1993).

The fashion sales of item i during the time interval t, denoted by FS_{it},can be represented as follows

$$FS_{it} = S_{it} + N_{it}, \tag{25}$$

where S_{it} and N_{it} denote the linear component and the nonlinear noise, respectively.

Assuming that $\tilde{S}_{it}$ is the forecasting value from the panel data, the nonlinear behaviour that the panel data can hardly capture is described as:

$$\lambda_{it} = FS_{it} - \tilde{S}. \tag{26}$$

Since the actual sales of fashion product will be affected by many non-linear factors, the nonlinear error model, which follows the state space representation, is shown in the following:

$$\lambda_{it} = f_{it}(\lambda_{it-1}, \omega_{it}) \qquad \text{Tansfer Function} \tag{27}$$

$$y_t = \lambda_{it} + \tilde{S}_{it} + \nu_{it} \qquad \text{Measurement Function} \tag{28}$$

where ω_{it} and ν_{it} follow Gaussian distribution $i.i.d.\ N(0, \sigma^2_{\varepsilon_{i,t}})$ and $N(0, \sum i)$, independently. In the measurement equation, y_t is the observation from each forecasting step.

For the PF model, based on (Zhong, et al. 2010), its analytical model is introduced as follows. First, assuming that the probability density function (PDF) of the initial state $p(\lambda_0)$ is known, the optimized state estimation is obtained by calculating the degree of confidence of $p(\lambda_t \mid y_{1:t})$ in different states, $y_{1:t}$ represents a set of observations from period 1 to t. Then, the conditional density $p(\lambda_t \mid y_{1:t})$ is recursively updated according to Equations (7) and (8).

$$p(\lambda_t \mid y_{1:t-1}) = \int p(\lambda_t \mid \lambda_{t-1}) p(\lambda_{t-1} \mid y_{1:t-1}) d\lambda \quad , \tag{29}$$

$$p(\lambda_t \mid y_{1:t}) = \frac{p(y_t \mid \lambda_t)\, p(\lambda_t \mid y_{1:t-1})}{p(\mathrm{y}_t \mid y_{1:t-1})} \quad , \tag{30}$$

where $y_{1:t-1}$ is defined as the history observation sequence with the random variables. The denominator $\int p(\lambda_t \mid \lambda_{t-1}) p(\lambda_{t-1} \mid y_{1:t-1}) d\lambda$ is a constant, which is available from the likelihood function and the statistical characteristic of the observation noise. PF provides an approximate solution for the discrete-time recursive updating of the posterior probability density function $p(\lambda_t \mid y_{1:t})$. Under PF, the posterior distribution of λ_t is approximated by a collection of weighted particles $\lambda = \{\lambda_t^n, w_t^n\}_{n=1}^N$. The posterior density can be calculated by

$$p(\lambda_t \mid \mathrm{y}_{1:t}) \approx \sum_{n=1}^{N} w_t^n \delta(\lambda_t - \lambda_t^n) \quad , \tag{31}$$

where δ is the delta-Dirac function and the weight w_t^n of each particle is updated according to

$$w_t^n \propto w_{t-1}^n \frac{p(\mathrm{y}_t \mid \lambda_t^n)\, p(\lambda_t^n \mid \lambda_{t-1}^n)}{q(\lambda_t^n \mid \lambda_{t-1}^n, y_t)} \quad , \tag{32}$$

The importance function $q(\cdot)$, known as a proposed conditional distribution, is important in the performance of PF (Gordon, et al. 1993). In general, the closer the importance function $q(\cdot)$ to the distribution of $\mathrm{p}(\cdot)$, the better the approximation is. The aim of choosing the optimal importance function is to minimize the variance of the true weights so that degeneracy problem is diminished in one way. The details of choosing the optimal importance function can be found in (Djuric, et al. 2003).

After introducing the panel data and the PF models above, we propose the PDPF forecasting model as follows. Notice that it is based on a two-stage (training and forecasting) structure as shown in Figure 5.1 and Figure 5.2. The algorithm of PDPF is illustrated as follows:

Training:

Step1. Initialization: set $t = 0$.

Step2. Input the historical data $S_{i,1\ldots k}, P_{i,1\ldots k} / CS_{j,1\ldots k}, P_{j,1\ldots k}$ and test the Panel Stationary and Hausman. Then, make panel forecasting model.

Step3. Forecasting $S_{i,k\ldots k+m} / CS_{j,k\ldots k+m}$ by using panel forecasting model.

Step4. Input the historical data $S_{i,k\ldots k+m} P_{i,k\ldots k+m} / CS_{i,k\ldots k+m}, P_{i,k\ldots k+m}$.

Step5. Regress the predicting error of panel forecasting model and initial the state space representation of the error.

Step6. Output the panel forecasting model and the state space representation of forecasting error.

Forecasting:

Step7. Input historical data $S_{i,1\ldots k+m} P_{i,1\ldots k+m} / CS_{i,k\ldots k+m}, P_{i,k\ldots k+m}$, panel forecasting model and forecasting error initialization.

Step8. Forecast the sale S_{it+1} / CS_{it+1}. Estimate the forecasting error λ_{it+1} according to Equations (7)-(10). Then, output the forecasting result FS_{it+1} / FCS_{it+1}.

Step9. Input the observation $\tilde{S}_{it+1} / C\tilde{S}_{it+1}$ and resample the particles according to the importance function $q(\cdot)$ (Gordon, et al. 1993).

Step10. Let $t = t + 1$, go to step 8.

Step11. The end.

Training

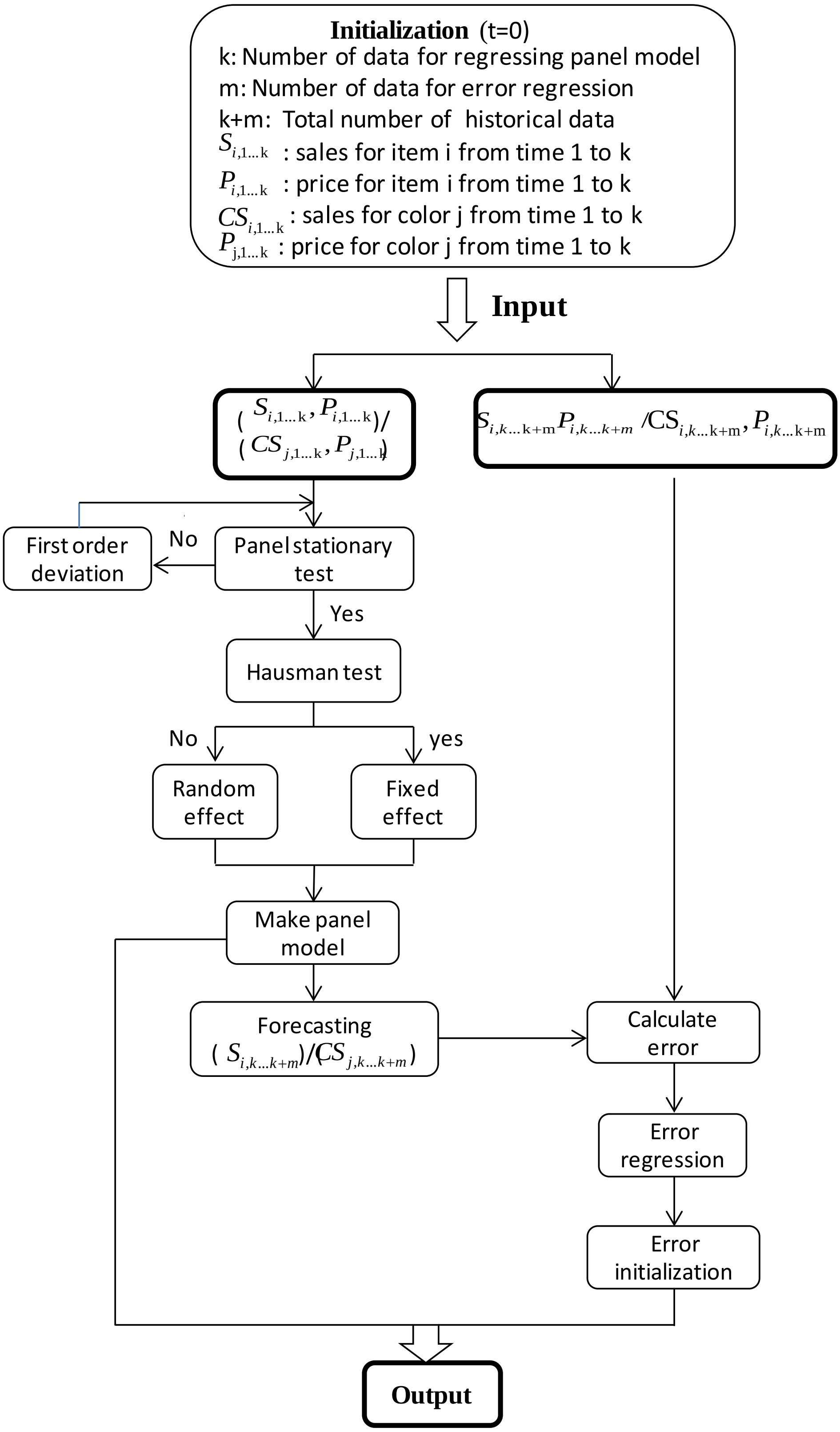


Figure 5.1 The training process for the PDPF forecasting model.

Forecasting

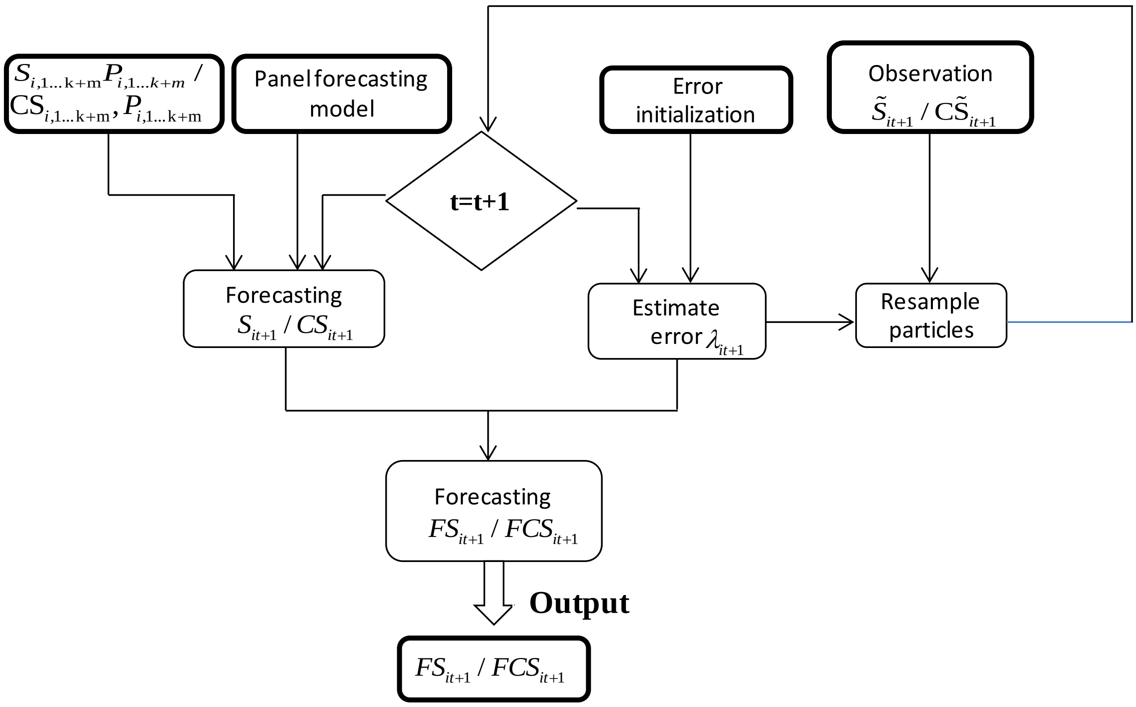


Figure 5.2 The PDPF forecasting process.

5.2 Case Study –Sales Forecasting by Item and Color

In order to show the forecasting accuracy performance of panel data based forecasting models, a linear statistic method and an intelligent method are taken as comparison methods in this chapter. Besides, to reveal more insights on the performance of our proposed model, we compare the forecasting performances with respect to two different kinds of fashion forecasting, namely the sales forecasting by item, and sales forecasting by color.

5.2.1 Datasets

In our analysis, in order to check the feasibility of our proposed model, we reclassify the weekly datasets according to item and color. In other words, the relationship between sales and price will be modeled in two types of data category, namely (i) item and (ii) color. The new datasets is illustrated in Table 5.1 and Table 5.1. After reclassification, the datasets of each category contain 36 samples. The first 24 samples are used as the training data (for estimating the model parameters) and the remaining 12 samples are used to do the forecasting test. Fashion products are featured as having a short life cycle (Choi et al. 2014). The historical sales data that can be used to conduct forecasting is limited. Thus, the research purpose of this study is to develop a forecasting model that can conduct fashion sales forecasting with a limited amount of historical data.

Table 5.1 Sale data and related price of different items (sample)

Number of week	Item	Quantity	Price
1	T-shirt	11	70.8
2	Dress	3	63.7

\

Table 5.2 Sale data and related price of different color (sample)

Number of week	Color	Quantity	Price
1	black	3	42
2	blue	11	76.7

5.2.3 Sales Forecasting by Items

To assess the forecasting performances of PDPF, the linear statistic forecasting method of Autoregressive Integrated Moving Average model (ARIMA) and the intelligent learning method of Extreme Learning Machine (ELM) are applied for conducting forecasting with the same datasets. ARIMA is a simple method based on the integration of auto-regression and moving average (Box, et al. 2013), which has been widely used in business forecasting. It is effective when the datasets is linear and with enough elements regressed and averaged. ELM has been known as a powerful tool for time series forecasting (Meng, et al. 2010). This new kind of kernel based approach, has been proven to be efficient in forecasting the nonlinear time series. As a comparison target, the pure panel data method, in which the PF part is not included, is also included in the study. The comparison of the overall forecasting errors for different items by using different methods is shown in Table 8. In the analysis, the sales amounts of different fashion items, including T-shirt, dress, bag, pants, accessory and belt are forecasted. MSE and SMAPE are used to assess the overall forecasting performances.

As shown in Table 5.3, in general, the intelligent method ELM does not perform better than ARIMA, panel data and PDPF. It is encouraging to note that the proposed PDPF model outperforms all the other methods including the pure panel data method in most cases. It is hence clear that the proposed method is suitable for conducting fashion sales forecasting. It is interesting to note that ARIMA, the commonly adopted linear regression method, which considers the items separately, cannot provide a good forecast performance because it neglects both nonlinear factors and items correlation. However, the products Bag and Belt are the exceptions with which the proposed PDPF model fails to provide better forecasting results than the other methods. A closer look into the product features reveals that the sales of Bag and Belt are likely to be seriously affected by the sales of other items (e.g., T-shirt, Dress, Pants), which could be explained mainly by the panel structure. The pure panel data method is hence very powerful and appropriate for forecasting the sales of these items. As a remark, in the real practice, there are a lot of fashions items need to be forecasted in order to make strategy decision. Traditional methods such as ARIMA and some artificial intelligence methods such as ELM might deliver good forecasting results for one or few items. However, the panel based method can forecast a lot of items simultaneously and deliver stable and accurate results in general.

Table 5.3 Forecasting comparison for different items by using different methods

Item	**ARIMA**		**ELM**		**Panel-data**		**PDPF**	
	MSE	SMAPE	MSE	SMAPE	MSE	SMAPE	MSE	SMAPE
T-shirt	62.75	9.77%	230.42	26.78%	64.45	10.60%	47.80	7.91%
Dress	3.37	29.82%	33.79	46.50%	3.02	27.36%	1.03	13.83%
Bag	18.07	44.94%	12.75	41.72%	1.00	17.44%	1.76	23.95%
Pants	19.55	16.11%	69.50	29.75%	13.12	10.76%	9.19	9.24%
Accessory	4.84	42.76%	31.40	52.48%	3.29	30.31%	1.11	18.99%
Belt	2.37	48.31%	52.22	70.59%	1.82	34.44%	5.58	52.03%

5.2.4 Sales Forecasting by Color

Sale forecasting by color is to estimate the sales quantity of different color. A panel relationship among sales in color category, the corresponding price and other influence factors will be constructed for the color forecasting. The testing process is similar to the sale forecasting model, and the result indicates that the color-price panel data is stationary and the fixed effect model would be selected. The forecasting model could be described as:

$$CS_{jt} = m + \alpha_j^* + \gamma S_{jt-1} + \beta \cdot \mathrm{P}_{jt} + \mu_{jt}, \quad j = 1,...,N, t = 1,...,T. \tag{33}$$

In this model, we use the total sales of each color to measure the color popular tendency, CS_j in the above equation; j represents different types of color which includes black, blue, brown, red, white, green, grey. The panel data estimation results are listed in Table 5.4.

Table 5.4 Estimation result of the color forecasting model

Coefficient	m	γ	β	α_1	α_2	α_3	α_4	α_5	α_6	α_7
Estimation	4.45	0.18	-0.0012	0.27	4.62	-1.39	-1.16	-0.82	1.14	-2.65
T-Statistic	4.63	2.20	-0.11	-	-	-	-	-	-	-

Similar to the estimation result of the SKU sale forecasting model, the fashion color trend also mainly depends on the whole market tendency and the previous color trend. However, the price effect is much less important in the color forecasting scenario than in the SKU sale forecasting scenario.

The sales forecasting accuracy under the item-based and color-based schemes is compared as follows. Firstly, Figure 5.3 and Figure 5.4 show the SMAPE of pure panel data method and the PDPF method for both forecasting schemes. The circle axis denotes the forecasting time points while the vertical axis refers to the error percentages. From Figure 4.4, it can be observed that, except for some special cases that SMAPE equals 100%, the forecasting accuracy of sales forecasting by item is more centralized and the mean SMAPE is better than that of sales forecasting by color. However, in Figure 5.4, the situation is totally different. In fact, the PDPF model improves the forecasting accuracy by 2.97% when the data is classified in color while the improvement of item category is only 0.83% in the same experimental situation. Compared to the pure panel data method, the PDPF model adopts PF to capture the nonlinear features that the panel data method cannot describe. From Figure 5.5, it is clear to note that the nonlinear error curve of the color category fluctuates more strongly than that of the item category, while the PDPF model can track this fluctuation trend better in the predicting process. Thus, the PDPF model performs better in dealing with the nonlinear forecasting problems.

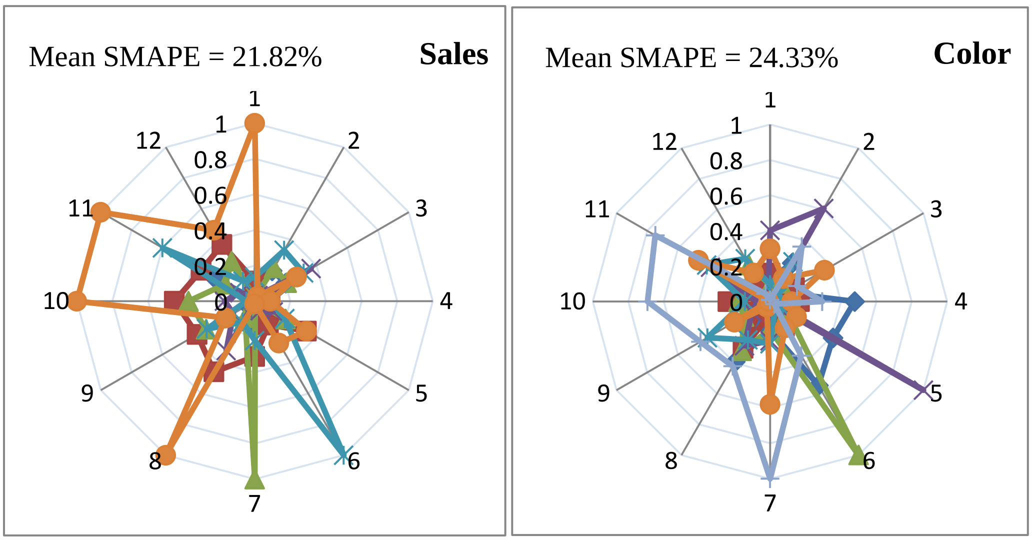


Figure 5.3 SMAPE results of pure panel data models for sales and color.

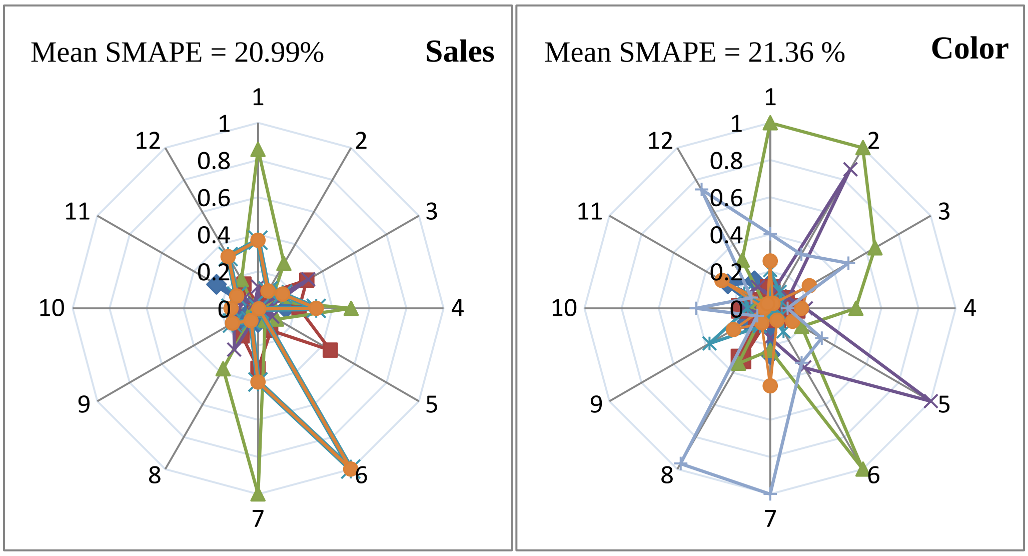
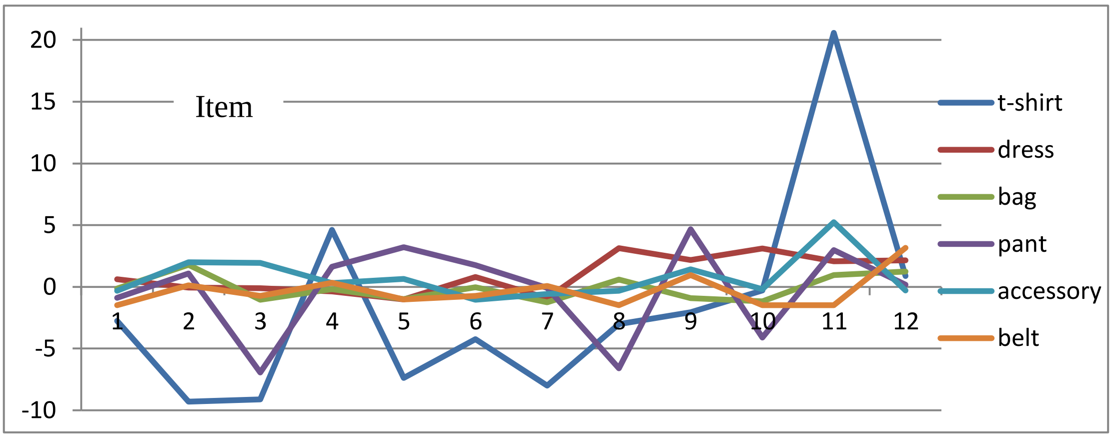


Figure 5.4 SMAPE results of the hybrid PDPF model for sales and color.



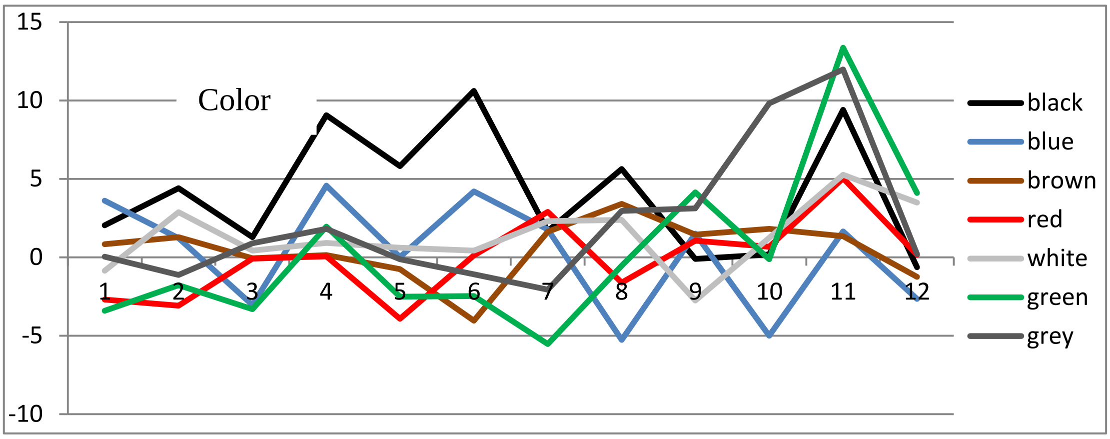


Figure 5.5 The forecasting error of the panel data model for sales quantity and color trend.

Chapter 6 Further Analysis for the PDPF Forecasting Model[6]

Fashion sales forecasting is always a challenging problem owing to many inherent features such as high volatility as well as limited data availability. To overcome the limitation of current forecasting methods based on time-series data, a panel data based forecasting model (PDPF) was proposed in Chapter 5. The core advantage of the proposed PDPF model is that the respective construct is a three-dimensional one which incorporates the time-series trend of previous sales, the price of product items under forecast, and the effects from other correlated product items, into the forecasting analysis. The comparison study indicated that the proposed model is applicable for conducting fashion forecasting and performs better than typical

[6] A part of this chapter was published in a journal paper: Ren, S., T.-M. Choi, and N. Liu. Fashion Sales Forecasting With a Panel Data-Based Particle-Filter Model. IEEE Transactions on Systems, Man, and Cybernetics: Systems 45(3), 411-421.

statistics model and AI model. In this chapter, how (i) the relationship between sales quantity and the corresponding price, (ii) the number of historical data and (iii) information updating and forecasting frequency, would influence the forecasting performance of PDPF would be further analyzed.

6.1 Effects of Degree of Correlation between Sales and Price

Since price is selected as a decision variable in the hybrid PDPF model, it is important to double check that whether the relationship between sales and the corresponding price would really influence the forecasting performance. In this section, the Granger Causality Test (Granger 1969) is adopted to measure this kind of potential relationship. Notice that y *is said to be Granger-caused by* x *if* x *helps in the prediction of* y*, or equivalently if the coefficients on the lagged* x*'s are statistically significant* (Granger 1969). It is important to note that, the statement "x *Granger causes* y" does not imply that y is the effect or the result of x. The testing result for causality only gives an indication of whether changes in prices are helpful in predicting changes in sales. The recently developed methodology in (Hurlin 2004) enables us to test for Granger causality in the panel data context. The causality test result and forecasting accuracy of our proposed PDPF model for each kind of item are listed in Table 6.1.

Table 6.1 clearly shows that except for T-shirt and Pants, all items fail to reject the null hypothesis of Granger Causality Test that "price does not Granger cause sale at the 5% confidence level". *In other words, the changes in price are "statistical-significantly" helpful in predicting the changes of sales for both T-shirt and Pants, but not the others*. This testing result reconfirms the fact that the forecasting results of PDPF for T-shirt and Pants outperform the other items (as shown in Table 6.1). The implication of this result is that one can apply the "Granger Causality Test" to examine the degree of correlation between price and sales quantity; *a larger degree of this correlation will imply a better forecasting result of our proposed PDPF model*.

Table 6.1 Granger causality test result of each item (null hypothesis: price does not granger cause sale)

Item	F-Statistic	Prob.	Test result	SMAPE
T-shirt	3.88	0.015	reject	7.91%
Dress	1.75	0.174	accept	13.83%
Bag	0.29	0.883	accept	23.95%
Pants	3.03	0.047	reject	9.24%
Accessory	1.38	0.271	accept	18.99%
Belt	1.41	0.261	accept	52.03%

*Significant at the 5% level.

6.2 Effects of Number of Historical Data

Sales forecasting in the fashion industry is known to be especially challenging because of the limited amount of data being available. This feature can be reflected by the fact that product lifecycle is short in fashion. As a result, a natural question on whether having fewer/limited historical data will lead to a worse forecasting performance compared to the case when historical data is abundant arises. In this sub-section, how the forecasting accuracy is related to the number of historical data for our proposed PDPF model would be examined. In order to make the comparisons, different sizes of historical data are selected for model training to conduct forecasting for the "future" 8 weeks demands correspondingly. The detailed results of different sizes of sample data are listed in Table 6.2. Specifically, the variation trend of forecasting accuracy for each item is explained in Figure 6.1. It can be observed that, the forecasting accuracy of these 7 items significantly fluctuates according to the number of historical data. Most items show a down trend when the number of historical changes from 9 to 28. A bit counter-intuitively, among them, the forecasting accuracies of T-shirt and Dress experience a decrease when the sample size of historical data grows up gradually. Thus, a larger set of historical training data fails to guarantee a better forecasting performance. In fact, for many cases, a medium amount of historical data yields the best result, and the forecasting performance difference between the cases having limited data and sufficient data is also relatively small. We can conclude that increasing of the amount of historical data does not necessarily improve forecasting accuracy under the proposed PDPF method. Furthermore, we argue that the PDPF method is applicable in the cases even with relatively limited amount of data which fits the fashion business operations well.

Table 6.2 SMAPE comparison of hybrid model by using different number of historical data

Item	9 (Min No.)	12	16	24	28
T-shirt	24.55%	30.24%	24.39%	15.98%	14.55%
Dress	112.38%	82.70%	68.63%	57.67%	22.20%
Bag	89.97%	56.42%	28.56%	51.93%	47.23%
Pants	39.46%	36.66%	25.68%	27.59%	16.46%
Accessory	84.94%	63.90%	66.41%	62.85%	75.52%
Belt	63.48%	56.21%	65.10%	52.41%	91.28%

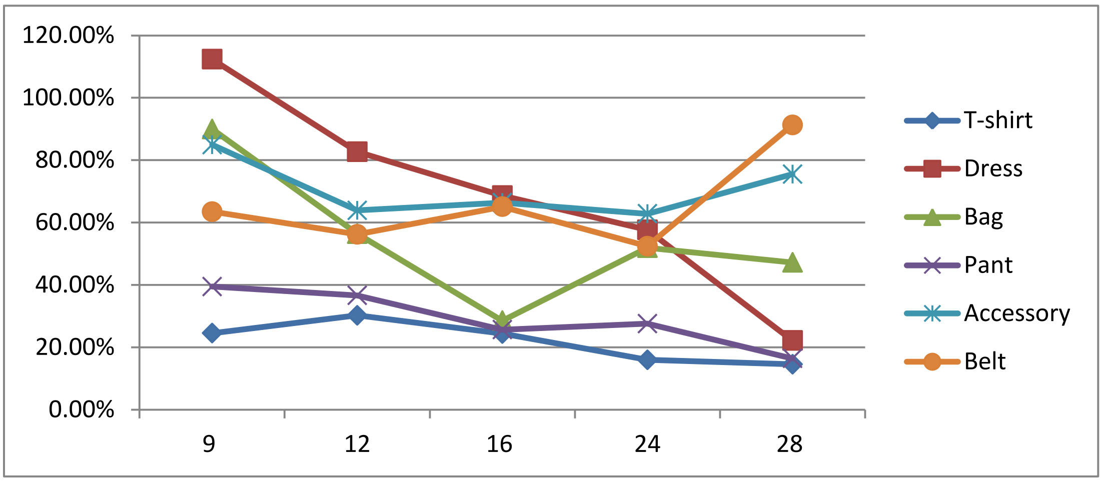


Figure 6.1 SMAPE of hybrid model by different number of historical data.

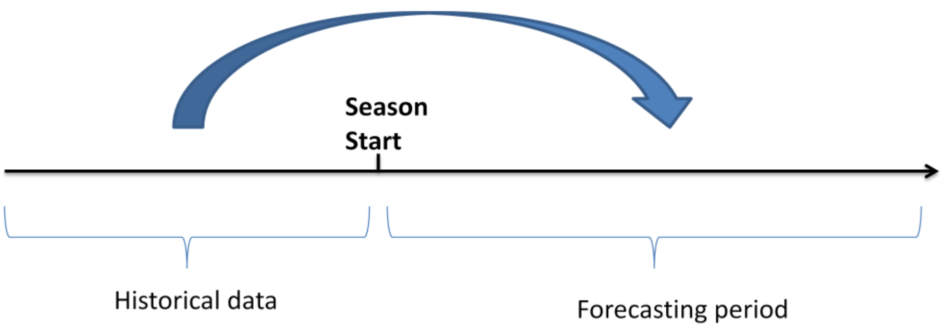
6.3 Effects of Information Updating and Forecasting Frequency

Short-time information updating is critically important for the sales forecasting during the selling season as it is well-known that recent sales information affects the future sales prediction significantly. However, in practice, it is nearly impossible to make forecast updating every day since there are a lot of SKUs and forecasting is time consuming. Therefore, exploring the proper forecasting and information updating frequency is very important. For this information updating process, a basic sales forecast is structured to predict the uncertain demand in the upcoming selling season first. Then, new sales data will be gained when the selling season starts. Pre-season forecasting results will be updated according to these new demand observations. In this section, the impact of the frequency of information updating on the forecasting performance of our proposed PDPF model would be examined.

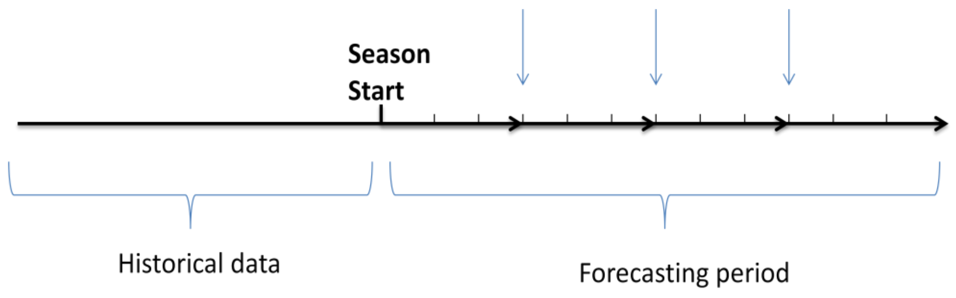
As described in Chapter 3, the PDPF forecasting model is based on two-component structure, namely the linear and non-linear components. The panel data method is used to investigate the relationship among forecasting impact factors in the linear component, while the nonlinear influence from the external environment such as the fashion trend, weather, holiday, etc. will be handled by the particle filter method. For the forecasting process, panel data model first makes a long-term forecast by regression. Then, particle filer updates the forecasting results in real time according to the latest information.

In our datasets, sales data of two fashion items (T-shirt and Pants) with 7 colors, together with other related properties of the items are included (c.f. Table 4.1). The dataset of each item contains 92 samples. The first 61 samples are used as the training data (for estimating the model parameters) and the remaining 31 samples are used to do the forecasting test. To access the impact of information updating frequency on forecasting performance, four forecasting scenarios were examined according to different updating frequency (i) no information update, (ii) with information updating every day, (iii) with information updating every two days and (vi) with information updating every three days. Those scenarios can be illustrated as below:

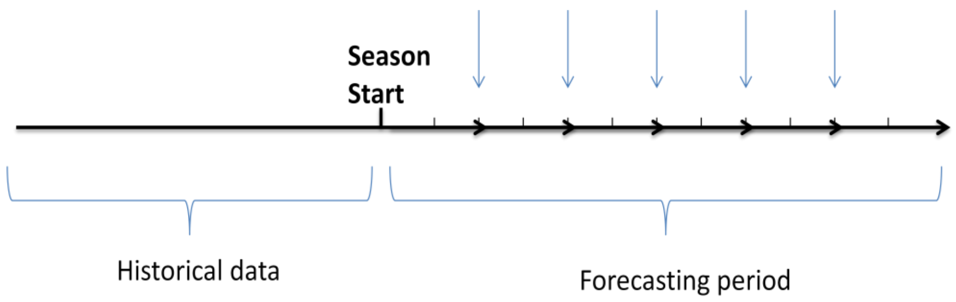
A. Forecasting without information updating



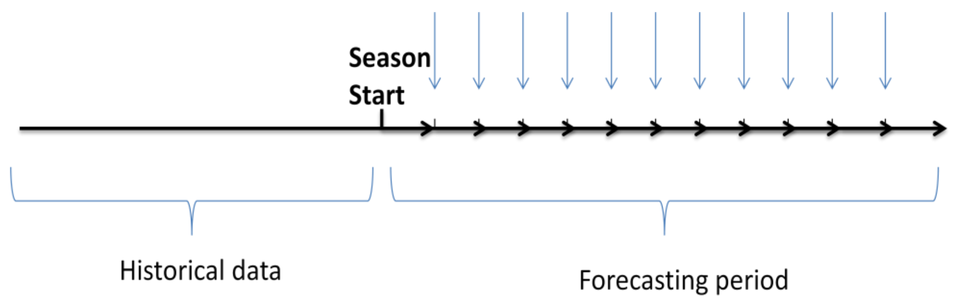
B. Forecasting with information updating every three days



C. Forecasting with information updating every two days



D. Forecasting with information updating every day



Observe that forecasting without information updating means using two-month historical data to predict the demand of the following one month by one time. While the forecasting process with information updating allows the forecasting results to be updated after any new sales data has been obtained. In order to explore the relationship between forecasting performance and the frequency of sales data updating, we choose three kinds of information updating frequency for simulation. In the results analysis, the mean absolute error (MAE) which is used to measure how close forecasts are to the eventual outcomes and SMAPE is used to measure the forecasting performance in a relative measure.

The forecasting results are shown in Table 6.3.

Table 6.3 Comparisons of forecasting results with respect to the information updating frequency

<table>
<tr><th colspan="2" rowspan="2">The frequency of information update</th><th colspan="2">T-shirt</th><th colspan="2">Pants</th></tr>
<tr><th>MAE</th><th>SMAPE</th><th>MAE</th><th>SMAPE</th></tr>
<tr><td colspan="2">Forecast without information updating</td><td>2.39</td><td>30.93%</td><td>1.46</td><td>37.31%</td></tr>
<tr><td rowspan="3">Forecast with information updating</td><td>update every three days</td><td>1.58</td><td>17.73%</td><td>1.37</td><td>30.22%</td></tr>
<tr><td>update every two days</td><td>1.45</td><td>14.28%</td><td>1.27</td><td>29.38%</td></tr>
<tr><td>update everyday</td><td>1.02</td><td>10.73%</td><td>0.95</td><td>22.43%</td></tr>
</table>

In general, forecasting with information updating outperforms forecasting without information updating for the sales forecast of T-shirt and Pants. Moreover, the frequency of information updating influences the forecasting performance; a higher frequency implies a better forecasting performance.

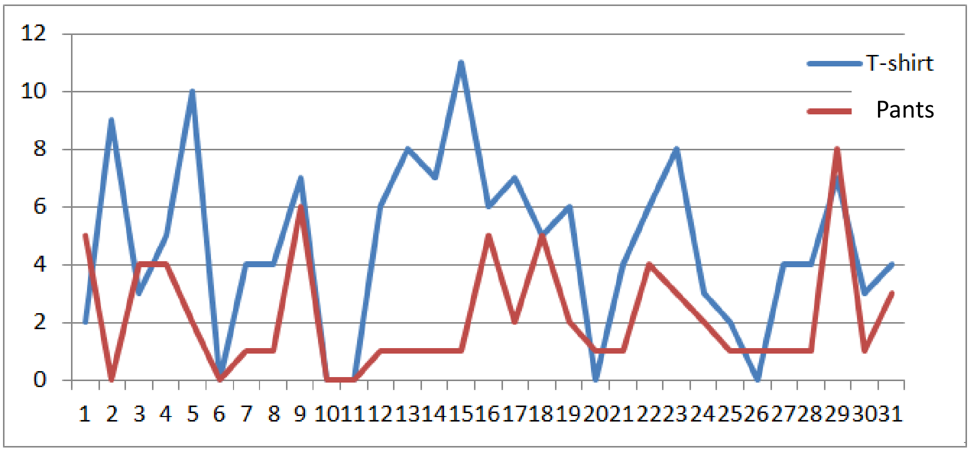


Figure 6.2 The real sales of T-shirt and pants.

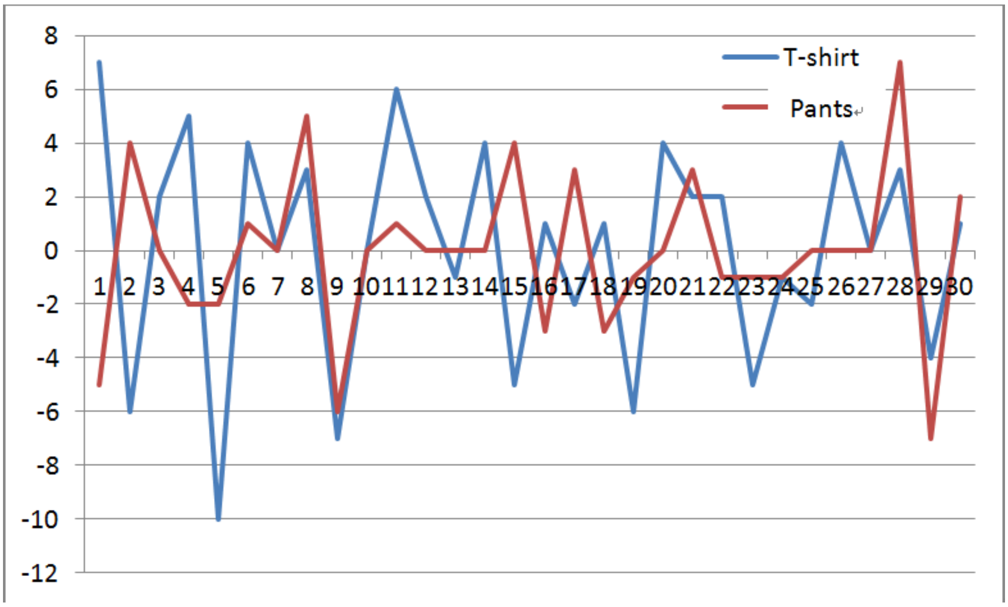


Figure 6.3 First order derivative of the real sales data of T-shirt and pants.

Next, how the pattern and feature of product sales relate to the forecasting performance with respect to the information updating frequency would be examined.

Figure 6.2 shows the real sales data of T-shirt and pants during our predicting period. To investigate the daily changing rate of the two curves, the first order deviative (FOD) of the real sales data of T-shirt and pants are obtained. The new series are shown in Figure 6.3. Then, *Standard Deviation* and *Skewness* are selected to test this changing rate. Observe that skewness is a measure of asymmetry of the distribution of the series around its mean. The testing results are listed in Table 6.4. It can be clearly seen that the daily changing rate of T-shirt is larger than that of pants. In other words, the sales of T-shirt fluctuate more than pants. It is interesting to note that the forecasting accuracy of the T-shirt is improved more than pants when the frequency of information updating is increased. This can be observed in Figure 6.4 and Figure 6.5. Thus, one can conclude that: (i) Forecasting with information updating performs better than that without information updating; (ii) When increasing the information updating frequency, a better forecasting performance can be obtained; (iii) The forecasting accuracy of the item with larger sales fluctuation will be improved more than the item with smaller sales fluctuation when the frequency of information updating increases.

Table 6.4 Standard deviation and skewness of T-shirt and pants' sales data

Criteria	T-shirt	Pants
Std. Dev.	4.18	3.05
Skewness	-0.05	-0.56

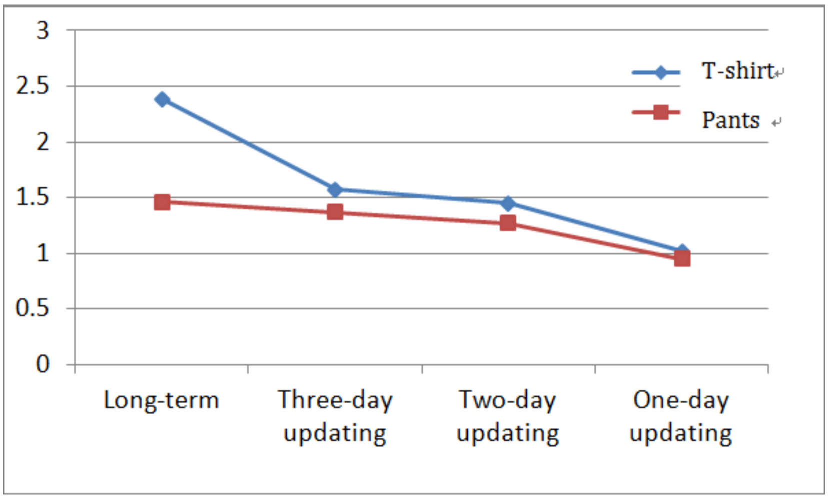


Figure 6.4 MAE of T-shirt and pants.

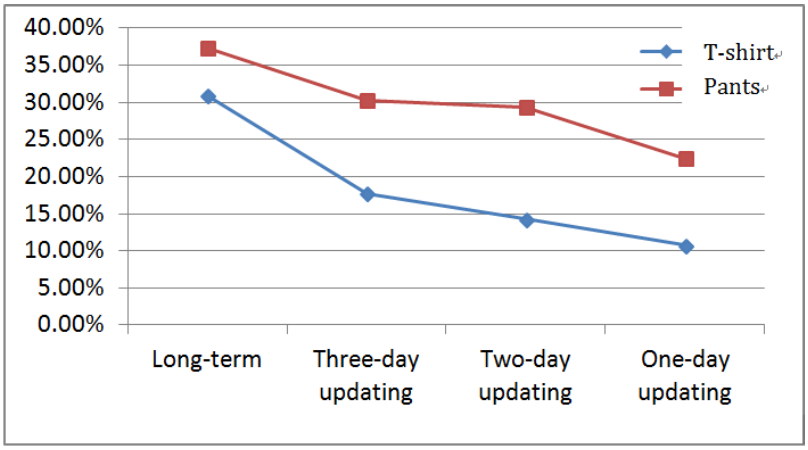


Figure 6.5 SMAPE of T-shirt and pants.

Chapter 7 Conclusion and Future Research

7.1 Conclusion

This thesis explored the fashion sales forecasting problems, with a focal point on the panel data based models. To be specific, the cutting edge technologies and methods of fashion sales forecasting were first examined in Chapter 2 via a comprehensive literature review. It was found that the pure time series statistical methods and the AI methods were commonly used but they all had deficiency. Thus, the panel data based models could be a good candidate for conducting fashion sales forecasting. Since the panel data based model seemed to be a promising tool, various important panel data based models were further investigated and examined in Chapter 3. The respective tests and estimations methods were also explored.

As fashion sales forecasting models should be useful and applicable for real world fashion businesses, in Chapter 4, an industrial survey was reported and the respective AHP analysis was conducted. To generate further insights on the preference of different decision makers with different roles in the fashion companies, further comparison studies were carried out with. Some important findings, including the usefulness of panel data based model and the other competing models, were obtained.

In Chapter 5 and Chapter 6, a novel panel data based model, the PDPF, was thoroughly explored. This PDPF model included two parts, namel the panel data (PD) part and the particle filter (PF) part. The core strength of this novel hybrid PDPF model was: (i) The ability to conduct fashion sales forecasting by examining both the linear and non-linear parts of the data, and (ii) its three-dimensional structure which could conducting fashion sales forecasting by considering the influence of the previous sales of the specific product item, the price of product items under forecast, and the effects brought by other correlated product items, in the same time period. With real world sales data, a computational analysis was conducted to further reveal the forecasting performance of PDPF. Comparisons with other commonly adopted fashion sales forecasting models were also done. It was found that the PDPF outperformed these methods. Furthermore, some critically important relationships, including the relationship between sales and price, the relationship between the amount of historical data on PDPF's forecasting performance, and the impacts brought by the frequency of information updating and forecasting, were all examined.

7.2 Limitations and Future Research

From the above analysis, it is known that panel data based methods are applicable for conducting fashion sales forecasting problems. However, the impact factors considered in the panel data in this study are limited due to the lack of related information in the dataset. In order to better describe the relationship between sales and different kinds of impact factors, more important factors, such as weather effects, should be considered and modeled in future studies.

Despite obtaining some important findings for fashion sales forecasting by using panel data based models, this thesis research could also be extended further in several directions. For example, since sales forecasting is just a part of the fashion company's operations, it will be interesting to investigate how the panel data forecasting techniques can be integrated with other operations such as inventory management and transportation control. In addition, even though it is intuitive that a better sales forecasting performance (e.g., higher forecasting accuracy) is beneficial to fashion companies, how much "more" business value the fashion companies would gain is largely unknown. Thus, it is crucial to examine in future research the business value of forecasting accuracy improvement for fashion companies.

Appendix A: The questionnaire for Chapter 6

Sales Forecasting Tool Survey

(Only for academic research use)

1. Sales forecasting is critical to inventory planning in fashion business. Suppose that you are asked to choose a forecasting tool (e.g. application software) to help you with weekly sales forecasting, please rank the importance of each performance measure in your mind.

   （‘1’represents very unimportant, while ‘5’ denotes very important. ‘3’ means neutral.）

	Very Unimportant ←			→ Very important	
a) Forecasting accuracy:	1	2	3	4	5
b) Forecasting Computational Speed (FCS):	1	2	3	4	5
c) Data Sufficiency Requirements (DSR):	1	2	3	4	5
d) Stability:	1	2	3	4	5
e) Ease of implementation:	1	2	3	4	5

Remarks:

For b): FCS means how long we can get the sales forecasting results by using the forecasting tool.

For c): DSR means whether a lot of historical data will be needed or not in order to conduct sales forecasting.

For d) Stability means whether the forecasting result is stable (fixed) or not (varying) when repeat the forecasting process.

2. What is your position in the company?

   □ Managerial/senior

   □ Operational/junior

   □ Others__________

3. Regarding the flagship product in your company, please select the product nature.

- □ Fashionable
- □ Basic

4. Regarding the flagship product in your company, please select the major market.
   - □ Lower-end market (i.e. mass market)
   - □ Higher-end market
5. Your gender:
   - □ Female
   - □ Male

References

Aburto, L., and R. Weber. 2007. Improved supply chain management based on hybrid demand forecasts, *Applied Soft Computing* 7(1), 136-144.

Aksoy, A., N. Ozturk, and E. Sucky. 2012. A decision support system for demand forecasting in the clothing industry, *International Journal of Clothing Science and Technology* 24(4), 221-236.

Alvarez, J., and M. Arellano. 2003. The time series and cross- section asymptotics of dynamic panel data estimators, *Econometrica* 71(4), 1121-1159.

Amemiya, T. 1985. *Advanced Econometrics*. Harvard University Press.

Anderson, T. W., and C. Hsiao. 1981. Estimation of dynamic models with error components, *Journal of the American statistical Association* 76(375), 598-606.

Anderson, T. W., and C. Hsiao. 1982. Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* 18(1), 47-82.

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Studies in operational regional science, 4.

Anselin, L. 1999. The future of spatial analysis in the social sciences, *Geographic Information Sciences* 5(2), 67-76.

Anselin, L. 2001. Rao's score test in spatial econometrics, *Journal of Statistical Planning and Inference* 97(1), 113-139.

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

Anselin, L., and S. Hudak. 1992. Spatial econometrics in practice: A review of software options, *Regional Science and Urban Economics* 22(3), 509-536.

Arbués, F., R. Barberán, and I. Villanúa. 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, Spain* 24, 30-31.

Arbués, F., R. Barberán, and I. Villanúa. 2004. Price impact on urban residential water demand: a dynamic panel data approach, *Water Resources Research* 40(11).

Arbués, F., M. Á. Garcıa-Valiñas, and R. Martınez-Espiñeira. 2003. Estimation of residential water demand: a state-of-the-art review, *The Journal of Socio-Economics* 32(1), 81-102.

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

Arellano, M., and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *The Review of Economic Studies* 58(2), 277-297.

Armstrong, J. S..1985. *Long-range Forecasting From Crystal Ball to Computer*, Wiley, New York.

Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp. 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Transactions on Signal Processing* 50(2), 174-188.

Au, K.-F., T.-M. Choi, and Y. Yu. 2008. Fashion retail forecasting by evolutionary neural networks, *International Journal of Production Economics* 114(2), 615-630.

Auffhammer, M., and R. Steinhauser. 2007. The future trajectory of us CO_2 emissions: the role of state vs. aggregate information, *Journal of Regional Science* 47(1), 47-61.

Babel, B., E. Bomsdorf, and R. Schmidt. 2008. Forecasting German mortality using panel data procedures, *Journal of Population Economics* 21(3), 541-555.

Bai, J., and S. Ng. 2004. A PANIC attack on unit roots and cointegration, *Econometrica* 72(4), 1127-1177.

Balestra, P., and M. Nerlove. 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* 34(3), 585-612.

Baltagi, B. 2008. *Econometric Analysis of Panel Data*. John Wiley & Sons.

Baltagi, B. H. 2008. Forecasting with panel data, *Journal of Forecasting* 27(2), 153-173.

Baltagi, B. H., G. Bresson, and A. Pirotte. 2002. Comparison of forecast performance for homogeneous, heterogeneous and shrinkage estimators: Some empirical evidence from US electricity and natural-gas consumption, *Economics Letters* 76(3), 375-382.

Baltagi, B. H., G. Bresson, and A. Pirotte. 2003. Fixed effects, random effects or Hausman–Taylor?: A pretest estimator, *Economics Letters* 79(3), 361-369.

Baltagi, B. H., G. Bresson, and A. Pirotte. 2012. Forecasting with spatial panel data, *Computational Statistics & Data Analysis* 56(11), 3381-3397.

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

Baltagi, B. H., and J. M. Griffin. 1997. Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline, *Journal of Econometrics* 77(2), 303-327.

Baltagi, B. H., S. Heun Song, B. Cheol Jung, and W. Koh. 2007. Testing for serial correlation, spatial autocorrelation and random effects using panel data, *Journal of Econometrics* 140(1), 5-51.

Baltagi, B. H., and C. Kao. 2001. Nonstationary panels, cointegration in panels and dynamic panels: A survey. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1808022

Baltagi, B. H., and D. Li. 2004. *Prediction in The Panel Data Model With Spatial Correlation*. Advances in spatial econometrics, 283-295.

Baltagi, B. H., and D. Li. 2006. Prediction in the panel data model with spatial correlation: the case of liquor, *Spatial Economic Analysis* 1(2), 175-185.

Baltagi, B. H., and Q. Li. 1991. A joint test for serial correlation and random individual effects, *Statistics & Probability Letters* 11(3), 277-280.

Baltagi, B. H., and Q. Li. 1992. Prediction in the one- way error component model with serial correlation, *Journal of Forecasting* 11(6), 561-567.

Baltagi, B. H., and Q. Li. 1995. Testing AR (1) against MA (1) disturbances in an error component model, *Journal of Econometrics* 68(1), 133-151.

Baltagi, B. H., S. H. Song, and W. Koh. 2003. Testing panel data regression models with spatial error correlation, *Journal of Econometrics* 117(1), 123-150.

Banerjee, A., J. Dolado, and R. Mestre. 1998. Error- correction mechanism tests for cointegration in a single- equation framework, *Journal of time series analysis* 19(3), 267-283.

Banica, L., D. Pirvu, and A. Hagiu. 2014. *Neural Networks Based Forecasting for Romanian Clothing Sector*. Intelligent fashion forecasting systems: models and applications, 161-194.

Bhargava, A., L. Franzini, and W. Narendranathan. 1982. Serial correlation and the fixed effects model, *The review of economic studies* 49(4), 533-549.

Blundell, R. W., Smith, R.J. 1991. Initial conditions and efficient estimation in dynamic panel data, *Annales dÕ Economie et de Statistique* 20(21), 109-123.

Bond, S. R. 2002. Dynamic panel data models: a guide to micro data methods and practice, *Portuguese economic journal* 1(2), 141-162.

Borenstein, M., L. V. Hedges, J. Higgins, and H. R. Rothstein. 2010. A basic introduction to fixed- effect and random- effects models for meta- analysis, *Research Synthesis Methods* 1(2), 97-111.

Box, G. E., G. M. Jenkins, and G. C. Reinsel. 2010. *Time Series Analysis: Forecasting and Control*. Journal of time series analysis 31(4), 303.

Breitung, J., and S. Das. 2005. Panel unit root tests under cross- sectional dependence, *Statistica Neerlandica* 59(4), 414-433.

Breitung, J., and S. Das. 2008. Testing for unit roots in panels with a factor structure, *Econometric Theory* 24(1), 88-108.

Breusch, T. S. 1978. Testing for autocorrelation in dynamic linear models*, *Australian Economic Papers* 17(31), 334-355.

Breusch, T. S., and A. R. Pagan. 1979. A simple test for heteroscedasticity and random coefficient variation, *Econometrica: Journal of the Econometric Society* 47(5), 1287-1294.

Breusch, T. S., and A. R. Pagan. 1980. The Lagrange multiplier test and its applications to model specification in econometrics, *The Review of Economic Studies* 47(1), 239-253.

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

Cachon, G. P., and R. Swinney. 2011. The value of fast fashion: Quick response, enhanced design, and strategic consumer behavior, *Management Science* 57(4), 778-795.

Caro, F., and J. Gallien. 2007. Dynamic assortment with demand learning for seasonal consumer goods, *Management Science* 53(2), 276-292.

Caro, F., and J. Gallien. 2010. Inventory management of a fast-fashion retail network, *Operations Research* 58(2), 257-273.

Chakir, R., and J. Le Gallo. 2013. Predicting land use allocation in France: A spatial panel data analysis, *Ecological Economics* 92, 114-125.

Chang, Y. 2002. Nonlinear IV unit root tests in panels with cross-sectional dependency, *Journal of Econometrics* 110(2), 261-292.

Chen, F., and T. Ou. 2009. Gray relation analysis and multilayer functional link network sales forecasting model for perishable food in convenience store, *Expert Systems with Applications* 36(3), 7054-7063.

Choi, I. 2001. Unit root tests for panel data, *Journal of International Money and Finance* 20(2), 249-272.

Choi, I., and T. K. Chue. 2007. Subsampling hypothesis tests for nonstationary panels with applications to exchange rates and stock prices, *Journal of applied econometrics* 22(2), 233-264.

Choi, T.-M. 2013. *Fast Fashion Systems: Theories and applications*. CRC Press.

Choi, T.-M., C.-L. Hui, N. Liu, S.-F. Ng, and Y. Yu. 2014. Fast fashion sales forecasting with limited data and time, *Decision Support Systems* 59, 84-92.

Choi, T.-M., C.-L. Hui, S.-F. Ng, and Y. Yu. 2012. Color trend forecasting of fashionable products with very few historical data, *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews* 42(6), 1003-1010.

Choi, T.-M., Y. Yu, and K.-F. Au. 2011. A hybrid SARIMA wavelet transform method for sales forecasting, *Decision Support Systems* 51(1), 130-140.

Chung, B.D., J. Li, T. Yao, C. Kwon, and T. L. Friesz. 2012. Demand learning and dynamic pricing under competition in a state-space framework, *IEEE Transactions on Engineering Management.* 59(2), 240-249.

Collopy, F., and J. S. Armstrong. 1992. Expert opinions about extrapolation and the mystery of the overlooked discontinuities, *International Journal of*

Forecasting 8(4), 575-582.

De Toni, A., and A. Meneghetti. 2000. The production planning process for a network of firms in the textile-apparel industry, *International Journal of Production Economics* 65(1), 17-32.

Deng, J. L. 1989. Introduction to Grey system theory, *The Journal of Grey System* 1(1), 1-24.

Djuric, P. M., J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez. 2003. Particle filtering, *IEEE Signal Processing Magazine* 20(5), 19-38.

Drukker, D. M. 2003. Testing for serial correlation in linear panel-data models, *Stata Journal* 3(2), 168-177.

Druska, V., and W. C. Horrace. 2004. Generalized moments estimation for spatial panel data: Indonesian rice farming, *American Journal of Agricultural Economics* 86(1), 185-198.

Egger, S., A. Ilie, Y. Fu, J. Chongsathien, D.-J. Kang, and M. E. Welland. 2005. Dynamic shadow mask technique: A universal tool for nanoscience, *Nano Letters* 5(1), 15-20.

El-Bakry, H. M., and N. Mastorakis. 2008. A new fast forecasting technique using high speed neural networks, *WSEAS Transactions on Signal Processing* 4(10), 573-595.

Elhorst, J. P. 2003. Specification and estimation of spatial panel data models, *International Regional Science Review* 26(3), 244-268.

Falk, M. 2010. A dynamic panel data analysis of snow depth and winter tourism, *Tourism Management* 31(6), 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.

Frank, C., A. Garg, L. Sztandera, and A. Raheja. 2003. Forecasting women's apparel sales using mathematical modeling, *International Journal of Clothing Science and Technology* 15(2), 107-125.

Franzese, R. J., and J. C. Hays. 2007. Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data, *Political Analysis* 15(2), 140-164.

Frazier, C., and K. M. Kockelman. 2005. Spatial econometric models for panel data: incorporating spatial and temporal data, *Transportation Research Record: Journal of the Transportation Research Board* 1902(1), 80-90.

Frees, E. W., and T. W. Miller. 2004. Sales forecasting using longitudinal data models, *International Journal of Forecasting* 20(1), 99-114.

Frondel, M., and C. Vance. 2010. Fixed, random, or something in between? A variant of Hausman's specification test for panel data estimators, *Economics Letters* 107(3), 327-329.

Gardner, W. 1960. Dynamic aspects of water availability to plants, *Soil Science* 89, 63-73.

Garín-Muñoz, T., and L. F. Montero-Martín. 2007. Tourism in the Balearic Islands: A dynamic model for international demand using panel data, *Tourism Management* 28(5), 1224-1235.

Gengenbach, C., F. C. Palm, and J.-P. Urbain. 2009. Panel unit root tests in the presence of cross-sectional dependencies: Comparison and implications for modelling, *Econometric Reviews* 29(2), 111-145.

George, B. 1994. *Time Series Analysis: Forecasting & Control*, Pearson education India.

Getis, A. 2007. Reflections on spatial autocorrelation, *Regional Science and Urban*

Economics 37(4), 491-496.

Ghemawat, P., J. L. Nueno, and M. Dailey. 2003. *ZARA: Fast fashion*, Harvard Business School Boston, MA.

Gholami, R., X. Guo, M. D. Higon, and S.-Y. T. Lee. 2009. Information and communications technology (ICT) international spillovers, *IEEE Transactions on Engineering Management.* 56(2), 329-340.

Godfrey, L. G. 1978. Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables, *Econometrica: Journal of the Econometric Society* 46(6), 1293-1301.

Godfrey, L. G. 1978. Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables, *Econometrica: Journal of the Econometric Society* 46(6), 1303-1310.

Gordon, N. J., D. J. Salmond, and A. F. Smith. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings F (Radar and Signal Processing)*140(2), 107-113.

Granger, C. W. 1969. Investigating causal relations by econometric models and cross-spectral methods, *Econometrica: Journal of the Econometric Society* 37(3), 424-438.

Guggenberger, P. 2010. The impact of a Hausman pretest on the size of a hypothesis test: The panel data case, *Journal of Econometrics* 156(2), 337-343.

Hamzaçebi, C., D. Akay, and F. Kutay. 2009. Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting, *Expert Systems with Applications* 36(2), 3839-3844.

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators, *Econometrica: Journal of the Econometric Society* 50(4), 1029-1054.

Hausman, J. A. 1978. Specification tests in econometrics, *Econometrica: Journal of the Econometric Society* 46(4), 1251-1271.

Hausman, J. A., and W. E. Taylor. 1981. Panel data and unobservable individual effects, *Econometrica: Journal of the Econometric Society* 49(6), 1377-1398.

Ho, D. K.-Y., and T.-M. Choi. 2014. *Collaborative Planning Forecasting Replenishment Schemes in Apparel Supply Chain Systems: Cases and Research Opportunities*, Intelligent Fashion Forecasting Systems: Models and Applications, 29-40.

Höglund, L. 1999. Household demand for water in Sweden with implications of a potential tax on water use, *Water Resources Research* 35(12), 3853-3863.

Holly, A. 1982. A remark on Hausman's specification test, *Econometrica: Journal of the Econometric Society* 50(3), 749-759.

Holtz-Eakin, D., W. Newey, and H. S. Rosen. 1988. Estimating vector autoregressions with panel data, *Econometrica: Journal of the Econometric Society* 56(6), 1371-1395.

Houthakker, H. S., P. K. Verleger, and D. P. Sheehan. 1974. Dynamic demand analyses for gasoline and residential electricity, *American Journal of Agricultural Economics* 56(2), 412-418.

Hsiao, C. 2003. *Analysis of Panel Data*, Cambridge university press.

Hsiao, C. 2007. Panel data analysis-advantages and challenges, *Test*. 16(1), 1-22.

Hsiao, C., M. Hashem Pesaran, and A. Kamil Tahmiscioglu. 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods, *Journal of Econometrics* 109(1), 107-150.

Hsu, L.-C., and C.-H. Wang. 2007. Forecasting the output of integrated circuit industry using a grey model improved by the Bayesian analysis, *Technological Forecasting and Social Change* 74(6), 843-853.

Huang, G.-B., Q.-Y. Zhu, and C.-K. Siew. 2006. Extreme learning machine: theory and applications, *Neurocomputing* 70(1), 489-501.

Hui, C.L., T.W. Lau, S.F. Ng, and C.C. Chan. 2005. Learning-based fuzzy colour prediction system for more efective apparel design, *International Journal of Clothing Science and Technology*17(5), 335–348.

Hurlin, C. 2004. Testing Granger causality in heterogeneous panel data models with fixed coefficients, *Document de recherche. LEO* 5.

Im, K. S., M. H. Pesaran, and Y. Shin. 2003. Testing for unit roots in heterogeneous panels, *Journal of Econometrics* 115(1), 53-74.

Issler, J. V., and L. R. Lima. 2009. A panel data approach to economic forecasting: The bias-corrected average forecast, *Journal of Econometrics* 152(2), 153-164.

Jain, D. C., and R. C. Rao. 1990. Effect of price on the demand for durables: Modeling, estimation, and findings, *Journal of Business & Economic Statistics* 8(2), 163-170.

Jang, M. J., and D. W. Shin. 2014. Tests for random time effects and spatial error correlation in panel regression models, *Statistics* 48(1), 101-120.

Kanniainen, J., S. J. Makinen, R. Piche, and A. Chakrabarti. 2011. Forecasting the Diffusion of Innovation: A Stochastic Bass Model With Log-Normal and Mean-Reverting Error Process, *IEEE Transactions on Engineering Management.* 58(2), 228-249.

Kaya, M., E. Yeşil, M. F. Dodurka, and S. Sıradağ. 2014. *Fuzzy Forecast Combining for Apparel Demand Forecasting*, Intelligent Fashion Forecasting Systems: Models and Applications, 123-146.

Kelejian, H. H., and D. P. Robinson. 1998. A suggested test for spatial autocorrelation and/or heteroskedasticity and corresponding Monte Carlo results, *Regional Science and Urban Economics* 28(4), 389-417.

Keller, W., and C. H. Shiue. 2007. The origin of spatial interaction, *Journal of Econometrics* 140(1), 304-332.

Kesavan, S., V. Gaur, and A. Raman. 2010. Do inventory and gross margin data improve sales forecasts for US public retailers?, *Management Science* 56(9), 1519-1533.

Ledesma-Rodríguez, F. J., M. Navarro-Ibanez, and J. V. Pérez-Rodríguez. 2001. Panel data and tourism: a case study of Tenerife, *Tourism Economics* 7(1), 75-88.

Lee, C.-C., and J.-D. Lee. 2010. A panel data analysis of the demand for total energy and electricity in OECD countries, *The Energy Journal* 31(1), 1-24.

Lee, L.-f., and J. Yu. 2010. Spatial panels: Random components vs. fixed effects.

Lehmann, E. L., and G. Casella. 1998. *Theory of Point Estimation*. Springer Texts in Statistics, USA.

Lei, M., and Z. Feng. 2012. A proposed grey model for short-term electricity price forecasting in competitive power markets, *International Journal of Electrical Power & Energy Systems* 43(1), 531-538.

LeSage, J., and R. K. Pace. 2010. *Introduction to spatial econometrics*, CRC press.

Levi, B. H. B. a. D. 1986. Estimating Dynamic Demand for Cigarettes Using Panel Data The Effects of Bootlegging, Taxation and Advertising Reconsidered, *The Review of Economics and Statistics* 68(1), 148-155.

Levin, A., C.-F. Lin, and C.-S. James Chu. 2002. Unit root tests in panel data: asymptotic and finite-sample properties, *Journal of Econometrics* 108(1), 1-24.

Li, J., T.-M. Choi, and T. E. Cheng. 2014. Mean variance analysis of fast fashion supply chains with returns policy, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 44(4), 422-434.

Li, Q., and C. Hsiao. 1998. Testing serial correlation in semiparametric panel data models, *Journal of Econometrics* 87(2), 207-237.

Li, W., and H. Xie. 2014. Geometrical Variable Weights Buffer GM (1, 1) Model and Its Application in Forecasting of China's Energy Consumption, *Journal of Applied Mathematics* 2014, 1-6.

Li, X., F. Kong, Y. Liu, and Y. Qin. 2011. *Applying GM (1, 1) Model in China's Apparel Export Forecasting*, 2011 Fourth International Symposium on Computational Intelligence and Design (ISCID), 2, 245-247.

Li, X., C. Yu, S. Ren, C. Chiu, and K. Meng. 2013. Day-ahead electricity price forecasting based on panel cointegration and particle filter, *Electric Power Systems Research* 95, 66-76.

Lin, Y.-H., and P.-C. Lee. 2007. Novel high-precision grey forecasting model, *Automation in construction* 16(6), 771-777.

Liu, G. 2004. Estimating energy demand elasticities for OECD countries. A dynamic panel data approach, *Research Department of Statistics Norway.* Available at: http://hdl.handle.net/11250/180417.

Liu, N., S. Ren, T.-M. Choi, C.-L. Hui, and S.-F. Ng. 2013. Sales forecasting for fashion retailing service industry: a review, *Mathematical Problems in Engineering*, 2013. http://dx.doi.org/10.1155/2013/738675.

Lohse, G., S. Bellman, and E. Johnson. 2000. Consumer buying behavior on the Internet: Findings from panel data, *Journal of Interactive Marketing* 14(1), 15-29.

Ma, L., and K. Khorasani. 2004. Facial expression recognition using constructive feedforward neural networks, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(3), 1588-1595.

Maddala, G. S., and S. Wu. 1999. A comparative study of unit root tests with panel data and a new simple test, *Oxford Bulletin of Economics and Statistics* 61(S1), 631-652.

Maddison, D.. 2006. Environmental Kuznets curves: A spatial econometric approach, *Journal of Environmental Economics and Management* 51, 218-230.

Meng, K., Z. Y. Dong, D. H. Wang, and K. P. Wong. 2010. A self-adaptive RBF neural network classifier for transformer fault analysis, *IEEE Transactions on Power Systems*, 25(3), 1350-1360.

Mengi, O. O. and I. H. Altas. 2011. A fuzzy decision making energy management system for a PV/Wind renewable energy system, *in Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications (INISTA '11)*, Istanbul, Turkey, 436–440.

Mishra, B. K., S. Raghunathan, and X. Yue. 2009. Demand forecast sharing in supply chains, *Production and Operations Management* 18, 152-166.

Mostard, J., R. Teunter, and R. De Koster. 2011. Forecasting demand for single-period products: A case study in the apparel industry, *European Journal of Operational Research* 211(1), 139-147.

Mundlak, Y. 1978. On the pooling of time series and cross section data, *Econometrica: Journal of the Econometric Society* 46(1), 69-85.

Nenni, M. E., L. Giustiniano, and L. Pirolo. 2013. Demand forecasting in the fashion industry: a review, *International Journal of Engineering Business Management* 5(37), 1-6.

Nerlove, M. 1967. Experimental evidence on the estimation of dynamic relations from a time series of cross sections, *Econometric Studies Quarterly* 18(3), 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* 39(2), 359-382.

Nerlove, M., and P. Balestra. 1996. *Formulation and Estimation of Econometric*

Models for Panel Data, advanced studies in theoretical and applied econometrics 3, 3-22.

Ni, Y., and F. Fan. 2011. A two-stage dynamic sales forecasting model for the fashion retail, *Expert Systems with Applications* 38(3), 1529-1536.

Ning, A., H. C. Lau, Y. Zhao, and T. Wong. 2009. Fulfillment of retailer demand by using the MDL-optimal neural network prediction and decision policy, *IEEE Transactions on Industrial Informatics,* 5(4), 495-506.

Olatubi, W. O., and Y. Zhang. 2003. A dynamic estimation of total energy demand for the Southern States, *The Review of Regional Studies* 33(2), 206-228.

Pai, P.-F., and C.-S. Lin. 2005. A hybrid ARIMA and support vector machines model in stock price forecasting, *Omega* 33(6), 497-505.

Pao, Y.H., G.H. Park and D.J. Sobajic. 1994. Learning and generalization characteristics of the random vector functional-link net, *Neurocomputing* 6(2), 163-180.

Phillips, P. C., and D. Sul. 2003. Dynamic panel estimation and homogeneity testing under cross section dependence*, *The Econometrics Journal* 6(1), 217-259.

Polebitski, A. S. and R. N. Palmer. 2009. Seasonal residential water demand forecasting for census tracts, *Journal of Water Resources Planning and Management,* 136, 27-36.

Quah, D. 1992. The relative importance of permanent and transitory components: identification and some theoretical bounds, *Econometrica: Journal of the Econometric Society,* 107-118.

Quah, D. 1994. Exploiting cross-section variation for unit root inference in dynamic data, *Economics Letters* 44(1), 9-19.

Quah, D. T. 1996. Empirics for economic growth and convergence, *European Economic Review* 40(6), 1353-1375.

Ramos, C. M., and P. M. Rodrigues. 2013. The importance of ICT for tourism demand: A dynamic panel data analysis, *Quantitative Methods in Tourism Economics*,97-111.

Ren, S., T.-M. Choi, and N. Liu. Fashion Sales Forecasting With a Panel Data-Based Particle-Filter Model. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 45(3), 411-421.

Roget, F. M., and X. A. Rodríguez González. 2006. Rural tourism demand in Galicia, Spain, *Tourism Economics* 12(1), 21-31.

Rong, H.-J., Y.-S. Ong, A.-H. Tan, and Z. Zhu. 2008. A fast pruned-extreme learning machine for classification problem, *Neurocomputing* 72(1), 359-366.

Saaty, T.L. 1980. *The Analytic Hierarchy Process*. McGra w-Hill, New York.

Saaty, T. L. 1990. How to make a decision: the analytic hierarchy process, *European Journal of Operational Research* 48(1), 9-26.

Sakai, M., J. Brown, and J. Mak. 2000. Population aging and Japanese international travel in the 21st century, *Journal of Travel Research* 38(3), 212-220.

Sargan, J. D., and A. Bhargava. 1983. Testing residuals from least squares regression for being generated by the Gaussian random walk, *Econometrica: Journal of the Econometric Society*51(1),153-174.

Sarkar, P. 2003. Sequential Monte Carlo Methods in Practice, *Technometrics* 45(1), 106-106.

Schultz, H. 1935. Interrelations of demand, price, and income, *The Journal of Political Economy* 43(4), 433-481.

Serel, D. A. 2013. Flexible procurement models for fast fashion retailers, *Fast Fashion Systems: Theories and Applications* 459.

Song, H., and G. Li. 2008. Tourism demand modelling and forecasting-A review of recent research, *Tourism management* 29(2), 203-220.

Song, H., P. Romilly, and X. Liu. 2000. An empirical study of outbound tourism demand in the UK, *Applied Economics* 32(5), 611-624.

Song, H., and S. F. Witt. 2000. Tourism demand modelling and forecasting: Modern econometric approaches, *Annals of Tourism Research* 28(4), 1078-1080.

Sun, Z., K.-F. Au, and T.-M. Choi. 2007. A neuro-fuzzy inference system through integration of fuzzy logic and extreme learning machines, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* 37(5), 1321-1331.

Sun, Z.-L., T.-M. Choi, K.-F. Au, and Y. Yu. 2008. Sales forecasting using extreme learning machine with applications in fashion retailing, *Decision Support Systems* 46(1), 411-419.

Swamy, P., and S. S. Arora. 1972. The exact finite sample properties of the estimators of coefficients in the error components regression models, *Econometrica: Journal of the Econometric Society* 40(2), 261-275.

Taylor, W. E. 1980. Small sample considerations in estimation from panel data, *Journal of Econometrics* 13(2), 203-223.

Telser, L. G. 1962. The demand for branded goods as estimated from consumer panel data, *The Review of Economics and Statistics*, 300-324.

Thomassey, S. 2014. Sales forecasting in apparel and fashion industry: a review, *Intelligent Fashion Forecasting Systems: Models and Applications*,9-27.

Thomassey, S., and A. Fiordaliso. 2006. A hybrid sales forecasting system based on clustering and decision trees, *Decision Support Systems* 42(1), 408-421.

Thomassey, S., and M. Happiette. 2007. A neural clustering and classification system for sales forecasting of new apparel items, *Applied Soft Computing* 7(4), 1177-1187.

Thomassey, S., M. Happiette, and J. M. Castelain. 2002a. An automatic textile sales forecast using fuzzy treatment of explanatory variables, *Journal of Textile and*

Apparel, Technology and Management 3(1).

Thomassey, S., M. Happiette, and J.-M. Castelain. 2005a. A global forecasting support system adapted to textile distribution, *International Journal of Production Economics* 96(1), 81-95.

Thomassey, S., M. Happiette, and J. M. Castelain. 2005b. A short and mean-term automatic forecasting system-application to textile logistics, *European Journal of Operational Research* 161(1), 275-284.

Thomassey, S., M. Happiette, N. Dewaele, and J. Castelain. 2002b. A short and mean term forecasting system adapted to textile items' sales, *Journal of The Textile Institute* 93(3), 95-104.

Vaidya, O. S., and S. Kumar. 2006. Analytic hierarchy process: An overview of applications, *European Journal of Operational Research* 169(1), 1-29.

Vroman, P., M. Happiette, and B. Rabenasolo. 1998. Fuzzy Adaptation of the Holt–Winter Model for Textile Sales-forecasting, *Journal of The Textile Institute* 89(1), 78-89.

Vroman, P., M. Happiette, and C. Vasseur. 2001. A hybrid neural model for mean-term sales forecasting of textile items, *Studies in Informatics and Control* 10(2), 149-168.

Wallace, T. D., and A. Hussain. 1969. The use of error components models in combining cross section with time series data, *Econometrica: Journal of the Econometric Society* 37(1), 55-72.

Wang, D. 2011. Robust data-driven modeling approach for real-time final product quality prediction in batch process operation, *IEEE Transactions on Industrial Informatics* 7(2), 371-377.

Wang, K., Q. Gou, L. Yang, and S. Shan. 2013. Coordination of a fast fashion supply chain with profit-loss sharing contract. In: T-M, Choi (Eds.), *Fast Fashion*

Systems: Theories and Applications, CRC Press Taylor & Francis Group, 477.

Wang, Z.-X. 2014. Nonlinear Grey Prediction Model with Convolution Integral NGMC and Its Application to the Forecasting of China's Industrial Emissions, *Journal of Applied Mathematics* 2014, 1-9.

Westerlund, J. 2007. Testing for error correction in panel data*, *Oxford Bulletin of Economics and Statistics* 69(6), 709-748. Article ID 580161.

Wong, W., and Z. Guo. 2010. A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm, *International Journal of Production Economics* 128(2), 614-624.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*, The MIT press.

Wooldridge, J. M. 2002. Inverse probability weighted M-estimators for sample selection, attrition, and stratification, *Portuguese economic journal* 1(2), 117-139.

Xia, M., and W. Wong. 2014. A seasonal discrete grey forecasting model for fashion retailing, *Knowledge-Based Systems* 5(7), 119-126.

Xia, M., Y. Zhang, L. Weng, and X. Ye. 2012. Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs, *Knowledge-Based Systems* 36, 253-259.

Yelland, P. M. and X. J. Dong. 2013. Forecasting demand for fashion goods: a hierarchical Bayesian approach, *in Handbook on Intelligent Fashion Forecasting Systems: Models and Applications.* 71-99.

Yesil, E., M. Kaya, and S. Siradag. 2012. Fuzzy forecast combiner design for fast fashion demand forecasting, *2012 International Symposium on Innovations in Intelligent Systems and Applications (INISTA).*1-5.

Yokuma, J. T., and J. S. Armstrong. 1995. Beyond accuracy: Comparison of criteria used to select forecasting methods, *International Journal of Forecasting* 11(4), 591-597.

Yoo. H. 1999. Short term load forecasting using a self-supervised adaptive neural network, *IEEE Transactions on Power Systems* 14(2), 779–784.

Yu, J., R. de Jong, and L.-f. Lee. 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large, *Journal of Econometrics* 146(1), 118-134.

Yu, J., R. de Jong, and L.-f. Lee. 2012. Estimation for spatial dynamic panel data with fixed effects: the case of spatial cointegration, *Journal of Econometrics* 167(1), 16-37.

Yu, Y., T.-M. Choi, and C.-L. Hui. 2011. An intelligent fast sales forecasting model for fashion products, *Expert Systems with Applications* 38(6), 7373-7379.

Yu, Y., T.-M. Choi, and C.-L. Hui. 2012. An intelligent quick prediction algorithm with applications in industrial control and loading problems, *IEEE Transactions on Automation Science and Engineering* 9(2), 276-287.

Yu, Y., C.-L. Hui, and T.-M. Choi. 2012. An empirical study of intelligent expert systems on forecasting of fashion color trend, *Expert Systems with Applications* 39(4), 4383-4389.

Zambelli, M. S., I. Luna, and S. Soares. 2009. Long-term hydropower scheduling based on deterministic nonlinear optimization and annual inflow forecasting models, *2009 IEEE Bucharest PowerTech, Bucharest.* 1-8.

Zampighi, L. M., C. L. Kavanau, and G. A. Zampighi. 2004. The Kohonen self-organizing map: a tool for the clustering and alignment of single particles imaged using random conical tilt, *Journal of Structural Biology* 146(3), 368–380.

Zhong, J., Y.-f. Fung, and M. Dai. 2010. A biologically inspired improvement strategy for particle filter: Ant colony optimization assisted particle filter, *International Journal of Control, Automation and Systems* 8(3), 519-526.

Zhu,Q.-Y., A.K.Qin, P.N.Suganthan, and G.-B.Huang. 2005. Evolutionary extreme learning machine, *Pattern Recognition* 38(10),1759–1763.