Is the Past the Best Forecast of the Future in a Data Rich Environment? The Case of China

Macquarie University

Master of Research thesis

Faculty of Business and Economics

Qin Zhang

December 31, 2014

Contents

List of Figures

List of Tables

List of Abbreviations

ACF Autocorrelation function

AIC Akaike information criterion

AR Autoregressive

ARIMA Autoregressive integrated moving average

CPI Consumer price index

DI Diffusion index

FARIMA Factor-augmented autoregressive integrated moving average

FAVAR Factor-augmented vector autoregressive

MSFE Mean square forecasting error

PCA Principle component analysis

PPI:industrial goods Producer price index:industrial goods

VAR Vector autoregressive

Abstract

GDP growth rate and inflation are two of the most critical issues facing China's economy. To improve the GDP growth rate and inflation forecasts in a data rich environment, this thesis studies forecasting of China's four leading macroeconomic variables using six models. These variables are the consumer price index (CPI), industrial production, electricity production, and producer price index:industrial goods. The three factor models used are: the diffusion index (DI), factor-augmented autoregressive integrated moving average (FARIMA) model and factor-augmented vector autoregressive (FAVAR) model. The three univariate time series models are: autoregressive model (AR), autoregressive integrated model and simple exponential smoothing model. The predictors are summarised using a small number of indexes constructed by principal component analysis and then are used to construct one-,three-, and six-month-ahead forecasts using 36 predictors from 1997 through 2014.

Compared to benchmark AR forecasts, the forecasting results of the factor models showed that the DI and the FARIMA model generally do not improve forecasting performances for CPI, industrial production, and production of electricity in one-,three-, and six-month-ahead. Rather, the FAVAR model yields significant improvements over the benchmark AR model except for CPI in one-month-ahead forecast. Another notable result is that the two wining models in three-, and six-month-ahead forecasts: exponential smoothing and factor-augmented VAR model essentially produce the naive forecasts except for PPI:industrial goods. This implies

that the benefits of using complicated forecasting models such as diffusion index or factor-augmented VAR model are minor; naive forecasts are sufficient to explain the predictable dynamics of the CPI, industrial production and electricity production in three-, and six-month ahead. Overall, this study provides interesting results on forecasting China's macroeconomy in a data rich environment.

Declaration

Statement of Candidate

I certify that the work in this thesis entitled "Is the Past the Best Forecast of the Future in a Data Rich Environment?" The Case of China has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged.

Qin Zhang

Student ID: 43124771

8 October 2014

Acknowledgements

This thesis is not purely a result of my own efforts as a large group of dedicated and amazing people have supported and aided me in its writing.

First and foremost I would like to thank my supervisors Dr Natalia Ponomareva and Dr Christopher Heaton who without which I could not have written this thesis. Their comments, insights, encouragement and patience helped me greatly to both to persevere and complete this thesis that I feel proud of.

I would also like to thank Associate Professor Tony Bryant for additional assistance throughout entire MRES program. He is so willing to take time of his busy schedules to help MRES team choosing topics and discussing economics in depth.

I am also grateful to my peers in the economics MRES group. I could not have picked a better group to go through two years with. Their friendship, encouragement and passion for economics truly brought out the best in me.

Lastly I would like to thank my parents and girlfriend. Even through my parents are in China, they have mentally encouraged me to do something that was truly challenging. I could not ask for more in a family and I am truly grateful to have such as fantastic and loving family. My girlfriend Ya Wen is always by my side supporting me through the depressed and amused time during this year.

Chapter 1

Introduction

> *Even if growth moderates, China is likely to become a high-income economy and the world's largest economy before 2030, notwithstanding the fact that its per-capita income would still be a fraction of the average in advanced economies. But two questions arise. Can Chinas growth rate still be among the highest in the world even if it slows from its current pace? And can it maintain this rapid growth with little disruption to the world, the environment, and the fabric of its own society?*
>
> -China 2030: *Building a Modern, Harmonious, and Creative High-Income Society*, the World Bank report 2012.

Economic forecasting is important for real-world decision-making. In this regard, forecasting inflation is fundamental, since expectations of inflation affect central bank decisions regarding the future path of monetary policy and, in turn, private sector consumption and investment decisions[1]. Likewise, forecasting China's economic performances is important due to the major role of the China's economy in

[1] By informing the public about likely trends in inflation, forecasts can influence expectations and can therefore serve as a nominal anchor for example in the wage bargaining process or for other nominally fixed contracts like housing rents or interest rates.

global content.

China's impressive economic performances have captured the world's attention in recent decades. However, questions and doubts have been raised about future China's economic activities. In particular the questions about whether the China's economy could have a hard landing or soft landing and concerns about property bubble are continuously debated. Yet, there is no unambiguous consensus as to future trends and movements concerning China's economic activities; indeed, this unambiguous consensus might be because there are no consummate and empirically testable forecasting models for China's economy. This is proved by the existence of only limited literature about modelling and forecasting China's economy, especially for post global financial crisis period. Therefore, having a reliable and empirically testable model for China's economy is an immediate priority to enable the China's government to plan a soft landing and intervene in the property bubble.

Moreover, having a reliable and empirically testable forecasting model for China's economy is also vitally important for Australia. According to the Australian Department of Foreign and Affairs and Trade (2014), China is by far the Australian largest export market with total value of 84,963 million Australian dollar exported to China, accounting for 28.1% of the total export value in the financial year 2012-2013. More importantly, the total value of good exported to China constituted approximately 5% of total real GDP in year 2012-2013, which is an important and indispensable source of Australian GDP. The global financial crisis hurt many countries including American and UK and these countries suffered from the downturn of property market, slow growth rate of GDP and high rate of unemployment. Australian, in comparison, did not experience negative impacts from global financial crisis; indeed, according to Australian Bureau of Statistics (2014) the Australia even had a real GDP growth of 5% per year over last the 5 year and this is surprising growth of the Australian GDP has been contributed to by the mining sector boom. The biggest of contribution of the mining boom in Australia is from exponential growth of China's economy. Therefore, the future trend as to whether the China's economy could have a hard landing is extremely important to Australian export sector and economy activities and forecasting China's economy is first necessary to predict this.

The objective of this study is to examine whether the large dimensional appropriate factor model is a sensible way of forecasting China's consumer price index (CPI), industrial production, production of electricity and producer price index:industrial goods (PPI:industrial goods). With recent advances in information technique and data mining, it is possible to access thousands of economic time series data which might be potentially useful for forecasting purpose. In this matter, it is of paramount importance to develop an appropriate model within a data-rich environment which allows analysts forecasting several leading China's macroeconomic variables using large number of predictors. The original motivations of this study are twofold: (1) although there is extensively amount of literature on modelling and forecasting China's macroeconomic, those that use large number of predictors are rather limited, and (2) the concern the scepticism about the unreliability of China's official data. Given that the possibility of China's statistical authority falsifying the official reported output is relatively higher than in Western countries, the statistical factor models ,in particular the factor-augmented vector autoregressive (FAVAR), are very appropriate to model and forecast the China's economy(Fernald et al., 2014).

Economic forecasting has a long tradition. Policy makers, central banks, the general public and academics have all been interested in producing accurate forecasts. However, traditional economic models, such as univariate time series and multivariate vector autoressive (VAR) models, are limited in the sense that they cannot accommodate large numbers of time series. Stock and Watson (2002b) began a new promising strand of forecasting literature in which they proposed a statistical factor model the so called diffusion index (DI) forecasting methodology. The DI allows analysts to forecast macroeconomic variables using large number of predictors and conduct the forecasts in a handful manner. Stock and Watson (2002b) confirmed the view that the use of a large number of data series significantly improves the forecasts of key macroeconomic variables significantly. Bernanke and Boivin (2003) also stated that "Factor models provide a methodology that allows us to remain agnostic about the structure of the economy." Therefore, factor modelling provides potential benefits to forecasting models by incorporating the structure of the economy into models.

The DI forecasting methodology was first introduced by Stock and Watson (1998, 1999, 2002b) in which they used the static factor forecasting model or more commonly known as the DI forecasting methodology to forecast UK's inflation and a several leading US macroeconomic variables. After that, the DI has been eagerly taken on board by extensive literature. For instance, Artis et al. (2005) used DI to forecast UK inflatinon; Schumacher and Dreger (2002) used a dynamic factor model to forecast German GDP growth rate; Forni et al. (2003) used 447 monthly macroeconomic time series to predict the main countries of Euro area industrial production and the consumer price index. Factor model has also been applied to other fields. Bernanke and Boivin (2003) applied FAVAR in estimating policy reaction functions for the Federal Reserve Board in a data- rich environment.

With regards to applying factor model to China. The first attempt was on Mehrotra and Sánchez-Fung (2008) who applied 15 models including the DI and the FAVAR to forecast China's inflation. Motivated by Mehrotra and Sánchez-Fung (2008), Lin and Wang (2013) applied three dimensional reduction techniques including principle component analysis (PCA) to construct three different types of factor-augmented AR models and then used them to forecast China's inflation for three-month, six-month and twelve-month ahead. Although they found that all three factor models outperform the benchmark autoregressive (AR) model, they studies have three limitations that needs further research[2]

[2]Three shortcoming are:(1) the within-sample forecasting horizons are fixed. The common forecasting literature, in contrast, conduct the forecasts in the way that augmenting the length of within-sample-forecasting by one month for each forecast;(2) the CPI series in they study is year-on-year based, which is not a perfect measurement for monthly change of CPI; and (3) the

This study aims to narrow the research gaps and limitations in Lin and Wang (2013). The focus of this study is on the comparison of forecasting performances of six models for four leading China's macroeconomic variables for one-month, three-month and six-month ahead[3]. This study collects 36 monthly macroeconomic time series that represent an exhaustive description of the China's economy including measures of government activity such as government revenue and government expenditure; real economic indicators such as industrial sales and production of electricity; financial indicators such as money supply and interest rate; and trade activity such as import and export.

Using relative mean square forecasting error (MSFE), this study made two important findings. Firstly, the DI and factor-augmented autoregressive integrated moving average (FARIMA) generally do not contribute substantial improvements over benchmark AR forecasts except for PPI:industrial goods series. The FAVAR forecasts, on the other hand, yield improvements with respect to benchmark AR forecasts except for the CPI in one-month-ahead. In some case, improvements over benchmark AR are substantial (at least 25% improvement by MSFE). This result implies that the predictable dynamics of four China's leading macroeconomic variables can be explained by the vector structure of lag factors and lag variables. Secondly, the naive forecasts can sufficiently explain the predictability of CPI, industrial production, and production of electricity series in three-month-ahead and six-month-ahead forecasting horizon. This is a rather surprising result because it essentially insinuates that the best approach to predict CPI, industrial production and production of electricity three-month and six-month-ahead is today's CPI, industrial production and production of electricity.

This study is divided into five chapters:

CPI is not seasonally adjusted which might be heavily impacted by China's New Year.

[3]Three univariate time series models are: (1) a well-defined AR model with the lag length of AR model is chosen by AIC, (2) well-defined ARIMA model which orders of AR term and MA term are chosen by AIC, (3) a simple exponential smoothing model. Three statistical factor models which factors are estimated by PCA and selected by Bai and Ng (2002)'s information criteria

Chapter 1 is the introductory part of the study and establishes the motivations and significances of the study.

Chapter 2 presents and discusses the relevant literature which underpins the factor forecasting models. The literature on econometric model of the China's economy is discussed in detail in order to find out the shortcomings of each paper and how to improve them. The literature on the application of factor model is reviewed extensively in order to see what previous research has concluded about whether factor model can generate accurate forecasts than competing models.

Chapter 3 lays out the theoretical review of factor models and alternative rival models. The focuses of this chapter are on underlying assumptions of factor models and the form of factor forecasting methodology. Reviewing the assumptions of factor models is important because it gives guidelines on how to transform data appropriately in order to determine a true representation of factors. The form of factor forecasting methodology provides forecasting equations that are used to construct out-of-sample forecasts. This chapter also provides evaluation criteria to examine the forecasting performances.

Chapter 4 discusses data in depth. This chapter begins with the motivations for using the production of electricity and PPI: industrial goods. This is necessary because the falsification of China's official reported output in much of the literature, for example Rawski and Mead (1998), Holz and Lin (2001b), and Nakamura et al. (2014). PPI: industrial goods provides a measure of rate of inflation from producer perspective. This chapter also presents a discussion on the process of transforming each series including screening outlier, and whether a series should take the difference or difference of logarithm and standardisation of transformed data.

Chapter 5 deal with the empirical work done for the study. This section begin how to construct forecasts and how to compare forecasting performance based on relative MSFE. The relative MSFE is computed relatively to the MSFE and MAFE of the univariate AR model (so the AR forecast has relative MSFE of 1.00). In depth discussion of empirical results is also presented in this chapter.

Chapter 6 concludes the study and discusses the future directions of research which could further enhance the forecasting accuracy of China's macroeconomic variables in a data rich environment.

Chapter 2

Review of the literature

2.1 Introduction

The empirical literature on modelling and forecasting China's leading macroeconomic variables using large number of predictors is rather limited compared to that of Western countries. As China's economy has experienced tremendous economic reform and record-high GDP growth in past decades, the reform in 1978 progressed the China's economy gradually from an agriculture based economy to a commerce and state-owned enterprises based economy. A so-called socialist market economic system was established which means that the price of the majority of products are demand driven and a small proportion of products are still supply driven (Adler, 2013). This rapid structural transformation makes forecasting China's economy more difficult than for Western economies. On top of that, the scepticism about the unreliability of China's official data also enhanced the difficulty of forecasting China's macroeconomic variables.

Section 2.2 reviews the literature concerning the modelling and forecasting of China's economy. To facilitate the need for forecasting and policy implications, it is necessary to define what is considered as an appropriate model for the China's economy. An appropriate model is guided by the main criteria that all behavioural equations should be economically meaningful, all parameter should be time invariant, and dummy variables should be used as rarely as possible Qin et al. (2007). These main criteria provide the foundation to foundation to review and appraise the literature.

Section 2.3 presents a literature review concerning on the application of the statistical factor model. With recent advances in information technology and statistical theories, it is an inevitable fact that data will be available for many more series over an increasingly long time span. As a result, there is increasingly a need for information to mimic economic relationship. However, conventional univariate and multivariate forecasting model such VAR models can not accommodate large numbers of predictors. The factor model summarises the large number of predictors into a relative few number of factors. As such, it is capable of exploiting the information in a large dimensional data while keeping the size of the forecasting model small.

2.2 Review of China's econometric models

Academic literature on modelling the China's economy has a long traditions on macro-econometric research and is of great interest to economists and academia alike. The first attempt to formulate a large, complicated and economic theory based model started in the early 1980s, when the first macro-econometric model was closely linked to the government project LINK (Qin et al., 2007). The model was large in size and based on annual data. Models built using quarterly series and following the dynamic specification approach were first experimented by the Institute of Quantitative and Technical Economics of CASS. However, their models are currently out of maintenance (Liu 2003). Existing dynamic specification of models are described by Zheng and Guo (2013), Qin et al. (2007), Bennett and Dixon (1996), Fernald et al. (2014), Lin and Wang (2013), Mehrotra and Sánchez-

Fung (2008), and many others.

The stated goal of econometric models is to model a state of economy using statistical models such as time series models. Chow (1985) developed a crude model of China's economy which consisted of a consumption function and an investment function to explain China's national account identity having a constant annual price from 1953 to 1982. The model Chow used was two-stage linear regression in which in the first stage he estimated a consumption function and an investment function then in the second stage he combined them into national account identity. His findings are two folds: (1) the data confirm Robert Halls version of the permanent income hypothesis and the accelerations investment principle in China and (2) the model serves as a crude model for further research. In addition, Chow (2010) reaffirmed that exactly the same model can still successfully explain China's annual data from 1978[1] to 2006 and the permanent income hypothesis and the accelerations investment principle theory are still valid even after the great cultural revolution. His findings explained China's national income identity in a proper way but two primary shortcoming are noteworthy. The first limitation is that the model serves no forecasting purpose. Secondly, although Chows two papers are forerunners in the history of modelling the China's economy, the ordinary least square entails economic data are cross-sectional and time uncorrelated; these are too strong assumptions in reality.

Based on key features of the Asian Development Bank, another macroeconometric model of the China's economy was conducted by Qin et al. (2007). The model comprised household income and consumption function, investment, government, trade, production, prices, money and the labour market. The authors used an equilibrium-correction model to formulate all the behavioural elements which reflected the essence of a transitional economy such as China. Stochastic simulations were performed to forecast a few key variables and empirical results showed the

[1]China implemented a policy called Reform and Opening which reformed Chinas economy from a pure central plan economic towards to a mixed of market and central plan economy. This year is important in the study of China's economy partly because it brought the availability of economic data

model is immensely useful to capture the stochastic variation of the China's economy. However, although Qin et al. (2007)'s study is invaluable to measure and forecast effects of stochastic shock on China's economy, they model can not be used to forecast future economic activity.

Zhou et al. (2013) predicted China's CPI, growth rate of industrial added value, exchange rate and money supply using the term structure of credit spreads. They used the traditional Svensson model with genetic algorithms to obtain the interest rate term structures of government bonds and corporate bonds and calculates credit spreads as their differences and then incorporated these term structure of credit sppeads to VAR model. They found VAR models can predict the changes of Chinas macroeconomics well, which indicates that the term structure of credit spreads contains information of future changes of macroeconomic variables. However, two major limitations are: (1) they had no benchmark model to compare with, and (2) they used a fixed with-in-sample forecasting period.

Mehrotra and Sánchez-Fung (2008) studied the forecasting performances of 15 alternative models for China's inflation. More precisely, the investigation of their study tackled the following questions: Can the forecasting of inflation in China benefit from using many predictors? Using 36 predictors that represent exhaustive description of China's economy, they found only those models (factor-augmented VAR and factor-augmented AR) with large number of predictors via PCA outperform benchmark AR forecasts. However, one of major shortcomings is that their paper was published in 2008-in that time the global financial crisis had not emerged. Therefore, there is a need to re-examine the performances of factor model using data after the global financial crisis[2].

Motivated by Mehrotra and Sánchez-Fung (2008), Lin and Wang (2013) predicted

[2]As will be discussed in Chapter 5 as well as in Conclusion chapter, my results uphold the superiority of factor-augmented VAR model but reverse the forecasting performances of DI. This reversion might be because DI produces extremely bad forecasts during global financial crisis which can not be compensated by satisfying performances in non-crisis period.

China's inflation using AR forecast, ARMA forecast, factor-augmented AR through PCA, sliced inverse regression and partial least square. They found the factor-augmented AR model with optimal number of principal component outperformed rival models in out-of-sample forecasting horizon (three-month, six-month, and twelve-month ahead). However, their study has three shortcomings that require further research. First, they used year-on-year CPI as a measure of monthly inflation rate, which is not very appropriate to truly measure the month-on-month change of inflation. Second, they did not seasonally adjusted CPI. As a results, unadjusted CPI includes the effect of China's New Year, making a high probability of over-fitting the AR model. Third, they used a single within-sample period to construct forecasts for three-month, six-month, and twelve-month ahead. In contrast, conventional forecasting literature such as Stock and Watson (2002b) and Stock and Watson (1999) augmented the within-sample period by one month to obtain next out-of-sample forecast.

Nevertheless, there two remarkable limitations arising from reviewing the literature on modelling and forecasting China's economy. First, these studies did not consider the quality of China's data. As mentioned in Fernald et al. (2014), one of the urgent challenges of forecasting Chinas economy is the weak quality of reported output and inflation figures-which even Vice Premier Li Keqiang questioned as unreliable (Fernald et al., 2014). Academia also questions about quality of the China's official data. Holz and Lin (2001a) argued that China's National Bureau of Statistics' 1998 revisions to industrial enterprise categorization caused massive confusion and misinterpretations. Rawski and Mead (1998) used information from cost surveys to derive new estimated of China's farm workforce and found that standard yearbook data for China's farm labor force massively overestimated the number of China's farm workers and such overestimated may easily surpass 10 millions.

Holz and Lin (2001b) study the reliability of the China's industrial statistics and found two major structural breaks in 1993 and 1998 along with numerous partial revisions have cased severe comparability in both time series and cross-sectional data. Holz (2003) reviewed and examined some of the most recent criticism of statistics on China's industrial value-added and Gross Domestic Product. He argued that even through some of famous economists such as Rawski (1976) and Chow (1986) [3] believed the reliability of China's official data, the margin of error in much of the published data is likely to be sufficiently large to allow the statistical authorities having a choice of final value from a relatively wide range of equally correct value.

In 2006, China's National Bureau of Statistics undertook a benchmark revision of national income and product accounts statistics based on the findings of the 2004 economic census. Holz (2008) studied the this benchmark revision in depth and found three doubtful implications. First, the 2004 economic census results validate the provincial aggregate output values but invalidate the centre's national ones which is not plausible. Second, economy-wide as well as sectoral nominal values were revised but real growth rates of some sectors remained unchanged, which implies that at least the real growth rates for secondary sector are erroneous. Third, the benchmark revision raises questions about the quality and meaning of a large body of official statistics, and it casts doubt on the professionalism and sincerity of Chinas statistical authority

Nakamura and Steinsson (2014) conducted an alternative estimates of China's growth and inflation using detailed information on China's household purchasing patterns. They argued that as households become richer, a smaller fraction of total expenditures are spent on necessities such as grain and a larger fraction on luxuries such as eating out. To test it, they use systematic discrepancies between cross-sectional and time-series Engel curves to construct alternative estimates of China's growth and inflation. The results showed that the official inflation was understated while the growth rate of consumption were overstated in 2000's.

[3] Rawski (1976) argued that "most foreign specialists now agree that statistical information published in China's sources provides a generally accurate and reliable foundation on which to base further investigations." and Chow (1986) judged that "by and large China's statistics officials are honest."

In additional to the weak quality of China's reported data, the second limitation is concerned about the rapid change of economic structure. The rapid pace of institutional and structural change in China motivates our focus on the data on the recent period. Such institutional and structural change makes the availability of data relatively short. To deal with weak quality data and the rapid change of economic structure, Fernald et al. (2014) suggested using the factor model, in particular FAVAR. The FAVAR is an adapted version of a factor model in which explanatory variables are firstly estimated through a large number of predictors and then a conventional VAR model is estimated using these factors. The FAVAR approach is particularly well suited to economic modelling when output and inflation are imperfectly observed-latent variables (Bernanke and Boivin, 2003). Bernanke and Boivin (2003) and Bernanke et al. (2004) also suggested that if economic time series data such as output and CPI are not directly observable or the quality of data are not reliable, the FAVAR approach leads to better empirical estimates than other models.

2.3 Applications of factor model

The initial works of factor analysis were conducted by Sargent et al. (1977) and John (1977) who introduced the dynamic factor approach to macroeconomics. Sargent et al. (1977) examined a small system and conclude that two dynamic factors can explain 80% or more of the variance of major economic variables, including the unemployment rate, industrial production growth, employment growth, and wholesale price inflation. John (1977) in his PhD thesis applied dynamic factor models to macroeconomic data and analysed these models in the frequency domain for a small number of variables. However, assumptions on John (1977) and Sargent et al. (1977) is too restrictive in the sense they imposes orthogonality on the idiosyncratic components which is not appropriate for macroeconomic data.

The improvements on statistical factor model have been conducted by Stock and Waston (1998,1999, 2002a,2002b) and Forni et al. (2000) which they allowed serial correlation and weakly cross-sectional correlation of idiosyncratic components. Following this, the use of factor models to forecast macroeconomic variables can be seen extensively studied in forecasting literature such as Stock and Watson (2006), Giannone et al. (2008), Breitung and Eickmeier (2006) and Reichlin (2002), den Reijer (2005), Schumacher (2007), Marcellino et al. (2003), Graff et al. (2004), Artis et al. (2005), Gupta and Kabundi (2010), Moser et al. (2007). Literature on using the factor model to analyse the effectiveness of monetary policy can be seen in the work of Bernanke et al. (2004), Fernald et al. (2014) and many others.

2.3.1 Literature for factor model forecasting

Stock and Watson (1998) developed a DI forecast methodology and applied it to forecast US industrial production and inflation. They used a balanced panel of 170 monthly macro time series variables covering 1960:01-1997:09 and an unbalanced panel in which these 170 series variables were augmented by 54 additional monthly series. Measured by relative mean square forecast errors in which they set MSFE of AR to be 1, they found that DI forecasts outperform benchmark AR forecasts. However, they pointed out that one of future research's emphasis related to usefulness of DI is that even if the factor model captured the joint behaviours of macroeconomic time series and demonstrates superiority over benchmark AR forecasts, there is no apparent mathematical theory that explains why forecast based on DI should outperform AR model or other specialized model that have been shown to be empirically useful.

Similarly, Stock and Watson (1999) used large number of predictors to forecast U.S. inflation. Starting through conventional Phillips curve forecasts model, they proposed two research interests: (1) is the traditional Philips curve stable over time?

And if the answer is no then is there any alternative version of Philips curve that provides better empirical forecasts performances than the conventional Philips curve? (2) Can we incorporate other economic activities data when forecasting inflation? To answer these two questions, they used monthly US macroeconomic data covering from 1959:1 to 1997:9 to compare. They found that firstly Phillips curves specified with alternative measures of real economic activity can provide forecasts with smaller mean squared errors than those from unemployment-based Phillips curves. For instance, incorporating housing starts, capacity utilisation, the rate of growth of manufacturing and trades sales produces forecasts that are generally more accurate than forecasts constructed from Phillips curves using the unemployment rate. Secondly, relying simply on the conventional Phillips curve might be a mistake; instead it is possible to improve conventional Phillips curve forecasts using the DI through PCA. The factor model benefits the forecasting performances through the inclusion of many predictors hence it provides more information than conventional Phillips curve does.

Moreover, Stock and Watson (2002b) conducted a model the so called DI which they used a large number of predictors (215 predictors) to forecast eight US important real economic variables in six-month-, twelve-month- and twenty-four-month-ahead. The forecasting process is separated into two steps. Step one involves the data dimension reduction process in which large dimensional data are summarized into relatively small factors through PCAs and in step two the estimated factors are used to forecast eight leading variables in the US economy. Compared by MSFE, they found that majority of DI forecasts outperform the benchmark AR forecasts. However, one of limitation in the Stock and Watson (1998,1999 and 2002b) is that they used Bayesian information criterion (BIC) to select the number of factors in DI model, which works well in simple forecasting models such as AR or autoregressive integrated moving average (ARIMA) but has not been proved its usefulness in large dimensional factor models.

Followed by Stock and Watsons applications of the DI to forecast leading macroeconomic variables in US and UK, applied applications of the factor model have been shifted extensively to other economies which provided both favourable and unfavourable evidence for the usefulness of the factor models. There has also been theoretical improvement with regards to selecting the number of factors. Bai and Ng (2002) developed a set of information criteria to determine the number of factors to be retained in factor model. These criteria add penalty terms to the minimised objective function[4] so that they ensure consistency, i.e. the true number of factors is selected with probability one when N and T diverge (Artis et al., 2005). Hallin and Liška (2007) developed information criteria to select a number of factors in the dynamic factor model as opposed to the static factor model by Bai and Ng (2002).

Various other empirical studies provided additional favourable evidence for the forecasting accuracy of the factors models. The Euro area is most popular destination, in terms of the number of studies that have been conducted to there. Artis et al. (2005) used the DI model with about 80 variables to explain UKs economic activity and concluded that DI model with as less as six factors can substantially improve the forecasts accuracy. However, the difference between Artis et al. (2005) and Stock and Watson (2002b) is that Artis et al. (2005) used Bai and Ng (2002)'s information criteria to select the number of factor in DI forecasting model whereas Stock and Watson (2002b) used BIC model which does not guarantee to have consistency in the static factor model.

Moser et al. (2007) compared the forecasting performances of three models: DI, VAR and ARIMA. They proposed twelve-month out-of-sample forecasts of Austrian diverse types of inflation measurement. By applying data from aggregate HICP as well as its five sub-indices (processed food, unprocessed food, energy, industrial goods and services), they found that factor models outperform ARIMA models in terms of forecasting accuracy. In addition, the factor model turned out to be more

[4]Details of these criteria are discussed in Chapter 3

accurate than VAR models, except for the processed food index, for which factor model and VAR models show a fairly similar performances. Their analysis also suggested that the predictive accuracy of aggregate inflation can be improved by combining subindices into factor models.

Schumacher (2007) used both static and dynamic factor models to forecast German GDP and found that, based on mean forecast square error, both static and dynamic factor models outperform the simple AR model and VAR. Similarly, Camacho and Sancho (2003) applied DI to Spain and they found: firstly, two factors can sufficiently explained the predictable dynamic of core inflation and Spanish economy activity; and secondly, the forecasting performances of DI outperform benchmark AR forecasts. den Reijer (2005) studied the application of DI to forecast Dutch GDP growth rate and found DI generated better forecasts than AR did. Forni et al. (2003) used 447 monthly macroeconomic time series to predict the main countries of Euro area industrial production and consumer price index. Following Forni et al. (2000)'s generalized dynamic factor model and Stock and Watson (2002b)'s DI forecasting model, they found both factor models outperformed univariate AR forecasts for inflation at one-, three-, six-, and twelve-month ahead and industrial production at one- and three-month ahead. Based on large panel time series, Cristadoro et al. (2005) used factor model to project monthly inflation in Euro area. They found forecasts of factor model outperformeded competing models such as AR model and random walk model.

There is also large amount of forecasting literature for factor model outside of Euro. Gupta and Kabundi (2011) examined the performances of DI forecasting model in Stock and Watson (2002b) to South Africa. In additional to univariate AR model, they compared the performances of DI with unrestricted VAR, Bayesian VARs (BVARs) and a typical New Keynesian Dynamic Stochastic General Equilibrium (NKDSGE) model. They found that a specific form of a factor model exists, whether based on Bayesian assumptions or incorporating both static and dynamic factors, which tends to outperform all other competing models.

Liu and Jansen (2007) proposed several types of general structural the factor model and use these models to forecast U.S macro economic variables. Unlike Stock and Watson (2002b), who estimated factors from full datasets which do not usually had a structural interpretation, the choice of factors in Liu and Jansen (2007) are more structurally meaningful in the sense that they estimated factors from subsets of information variables, where these variables can be assigned to subsets on the basis of economic theory. They compared the forecasting performances of the various types of general structural components in the factor model with that of an AR, a standard VAR and some non-structural factor forecasting model. The analysis of results suggested that the accuracy and performances generated by the structural factor model are significantly better than competing models, especially at short horizons such as six- and twelve-month ahead forecasts.

Brisson et al. (2003) examines the usefulness of DI model using data of growth rate of real output and real investment in Canada. They compared forecast performances of DI to a variety of alternatives, including benchmark AR model and the forecasts made by the OECD. They found forecasting accuracy can be gained by using DI at short horizons but did not find such evidences for long horizons. Gosselin and Tkacz (2001) also studied the application of DI on forecasting Canada inflation. Using similar techniques to PCA, they extracted factors from a sample dataset consisting both Canadian and US macroeconomic data and found DI on average produced 30% less MSFE than that of AR forecast. Therefore, factor models are as good as more elaborate models in in forecasting Canadian inflation.

Giannone and Matheson (2007) introduced a new indicator of core inflation for New Zealand which is estimated through a dynamic factor model. They found this new indicator of core inflation provides relatively good forecasting performances when compared to a range of competing other models. Likewise, Graff et al. (2004) in the conference of Reserve Bank of New Zealand presented that although the primary source of formulating monetary policy and forecasting inflation by Reserve Bank of New Zealand is the Reserve Bank macro model, the leading indicators constructed through PCA is at least as useful and informative as the more preferred structural-based approaches to forecasting inflation in New Zealand.

2.3.2 Literature against forecasting ability of factor model

Just as a coin has two sides, there is literature against the forecasting ability of the factor model. Banerjee et al. (2005) evaluated the forecasting performances of many single indicators from European and US, factors extracted from a set of indicators, and groups of factors for inflation and GDP growth rate in Euro area. They found that static factor models - factors extracted from a set of indicators through PCA - do not always outperform single indicator methods and the best forecasting methods change over time. Schumacher and Dreger (2002) predicted German GDP using two classes of factor model: static factor model and dynamic factor model. In order to compare the performances of forecasting models, they presented a table that provides information on the ranking of forecasting performance of each model.

Table 2.1: Ranking of forecasting models by relative mean square forecasting error

	1	2	3	4	5	6	7	8
Static factor model	4	4	4	3	3	3	4	4
Dynamic factor model	1	1	1	1	1	1	1	1
ifo climate	2	2	2	2	2	2	2	2
VAR	3	3	3	4	4	4	3	3

Note: the numbers 1 to 8 in column reflect the forecasting horizon. For instance, 5 means five-month ahead forecast

A first glance at Table 2.1 reveals that the dynamic factor model produced the best forecasts among four competing models, but they used a test for equal forecasting accuracy which tested the improvements of dynamic factor model based on forecasting combing regression as below

$$y_t = (1 - \lambda)y_{A,t|t-h} + \lambda_{yB,t|t-h} + \varepsilon_t \tag{2.1}$$

They concluded that such improvements are statistically insignificant; therefore, the efficiency gains of using a large data set with dynamic factor models seem to be limited.

In order to examine the inflation forecasting performances of extracted factors at the aggregate Euro area level, Angelini et al. (2001) applied DI to forecast Euro-wide inflation using a multi-country data set and a broad array of variables. Their analysis of results were vague. On one hand, they have concluded that the factors extracted from aggregate Euro dataset through PCA are the most relevant for inflation forecasting in the with-in-sample forecasting horizon. On the other hand, they partially reversed the finding in the within-sample forecasting horizon and have concluded that the factors are not relevant for inflation forecasting in the out-of-sample forecasting period and the efficiency of the factors is blurred.

Schumacher and Breitung (2008) used the DI to forecast German GDP with monthly and quarterly data. To deal with mixed frequency of data and missing observations, they employed an expectation-maximisation algorithm and PCA to extract factors. Although their analysis of results suggested the dynamic factor model can generate superior forecasts than AR model, these improvement are moderate or even minor. Likewise, the study conducted by DAgostino et al. (2006) showed the improvements of the factor model were unclear. They predicted diverse measures of inflation: consumer price index, producer price index and personal consumption expenditure deflator using naive forecast, univariate AR model, factor-augmented AR model and pooling of bivariate forecasts. They found that, even though the vast majority of factor-augmented AR models outperform the benchmark naive model, they were not necessary to outperform univariate AR model for three measures of inflation.

Therefore they concluded that the simple AR model can hardly be outperformed by more sophisticated models.

2.4 Conclusion

This chapter has been focused on reviewing the empirical literature concerning economic models of China's economy as well as both literature for and against factor model.

Section 2.2 reviewed and examined the existing literature on the appropriateness of models of China's economy, some of which focused on using econometric models such as error correction model in Qin et al. (2007) or time series model such as Lin and Wang (2013) and Mehrotra and Sánchez-Fung (2008). Two of the major shortcomings in these studies are concerned with weak quality of data and rapid chnage of economic structure. This makes the factor model, especially FAVAR model, very appropriate (Bernanke et al., 2004; Fernald et al., 2014). In additional, because of the ability of summarising all the relevant information in China which which includes demand-oriented behaviour such as the price of commodities and retail sales as well as supply-driven component such as production of oil, the factor model is able to reflect the essence of the unique structure of the China's economy[5].

Section 2.3 reviewed the relevant literature on the effectiveness of the factor forecasting model and found there is limited consensus as to what makes forecasting performances effective. Despite that, there is overwhelmingly more evidence to support the fact factor forecasting model contributing improvements to forecasting accuracy and some literature such as Stock and Watson (2002b) and Forni et al. (2003) found these improvement can be substantial, there is also literature against the effectiveness and usefulness of the factor model like Angelini et al. (2001), DAgostino et al.

[5]the vast majority of components in the model are demand-oriented to reflect a high-degree of marketisation and it also has a number of supply-driven components.

(2006) and Schumacher and Breitung (2008). The conclusion that has the most empirical literature support is that factor model is better tool to forecast Euro and U.S. leading macroeconomic variables than AR model.

Chapter 3

Methodologies of forecasting

3.1 Introduction

This chapter lays out the theoretical underpinnings of the AR and the factor models used for forecasting China's leading macroeconomic variables multi-periods ahead. In order to evaluate the performances of forecasts, one must have a competing model or a benchmark model. A well defined order of AR model is suggested by many forecasting literature such as Stock and Watson (2002b) and Reichlin (2002) as the most common benchmark model to compare forecasts accuracy among various rival models. Motivated by Lin and Wang (2013), this paper also considers two classes of traditional univariate time series model: ARIMA model and simple exponential smoothing model.

The factor models used are motivated by Stock and Watson (2002b) and Mehrotra and Sánchez-Fung (2008) in which they used DI and FAVAR model to forecast US and China's leading variables. To examine the effectiveness of factor models further, this study also augments factors to a standard ARIMA and VAR to form a FARIM and FAVAR forecasting methodology. Suggested by Stock and Watson

(2002b) and Bai and Ng (2008), this study uses PCA to estimate factors because when N is sufficiently large then they are precisely adequate to be the proxy of data in subsequent forecasting models.

3.2 Theoretical underpinnings of factor model

This section surveys the theoretical works on the three statistical factor models: the DI, the FARIMA and the FAVAR model. The factor model has received considerable attention in the past decades because of its ability to appropriately and consistently forecast macroeconomic variables under a data-rich environment by reducing large amount of information or predictors to a manageable size. Theoretically speaking, the premise of the factor model is that one can admit there exist few latent and unobservable factors - f_t driving the co-movements of high-dimensional vector of time series, X_t, which is also affected by a by a vector of mean-zero idiosyncratic disturbances, e_t. In practice, the high-dimensional vector of time series X_t could be a state of economy, in the case of this thesis, the China's economy. As for mean-zero idiosyncratic disturbances, the e_t could be, for example, measurement error from data collection process. Therefore, the idea of using the factor model to forecast China's macroeconomic variables arises from two parts. The first part is associated with diverse range but short history data from China's Bureau Statistic[1]. The second part is the premise of the factor model in which one can estimate latent and unobservable factors first which contains a large amount of information from a great number of predictors and then use some of these factors to forecast China's macroeconomic variables.

To understand the theoretical underpinnings of the factor model, we begin by a discussion of the statistical model that motivates the DI. Let N be the number

[1]as mentioned before,there are over 500 predictors available in China's Bureau Statistic but for the majority of predictors the availability of time horizons is less than 20 years

of variables and T be the number of time series observations. Following Stock and Watson (2005), we suppose that (X_t, F_{t+1}) admits a dynamic factor model representation with:

$$X_t = \Lambda F_t + D(L)X_{t-1} + \nu_t \tag{3.1}$$

$$F_t = \Theta(L)F_{t-1} + \eta, \tag{3.2}$$

where X_t is an $n \times 1$ vector of stationary economic variables, F_t is a $r \times 1$ vector of unobserved common factors, with $r < n$. The equation (3.1) and (3.2) is known as the dynamic factor model in statistical literature. Stock and Watson (2002b) imposed two important modifications to the dynamic factor model. First, the lag polynomials $\lambda_i(L)f_t$, $\beta(L)f_t$, and $\gamma(L)y_t$ are restricted to have finite orders of maximum q orders[2]. The finite lag assumption allow the dynamic factor model to be written as a static factor model and rewrite as:

$$y_{t+1} = \beta' F_t + \gamma(L)y_t + \epsilon_{t+1} \tag{3.3}$$

$$X_t = \Lambda F_t + e_t, \tag{3.4}$$

where F_t=(f_t', f_{t-1}',.........,$ft - q'$)' is r×1 ($r \leq (q+1)\bar{r}$), the i^{th} row of Λ in equation(9) is (λ_{i0}, λ_{i1}.........,λ_{iq}), and β =(β_0,β_1........,β_q)'.

Although the knowledge of the dynamic representation of factor model is useful to some extent such as in precisely establishing the number of primitive shocks in the economy, the static representation of factor model is much better to understand in terms of theoretical standpoint and many econometric frameworks can be developed within a static framework (Bai and Ng, 2008). From a practical perspective, the primary benefit of rewriting an approximate dynamic factor model as a static representation is that latent and unobservable factors can be estimated by time domain methods such as PCA (Bai and Ng, 2008). It also involves few choices of auxiliary parameters. On the other hand, the estimation of the dynamic factor model entails

[2]so that $\lambda_i(L)$=$\sum_{j=0}^{q} \lambda_{ij}L^j$ and $\beta(L) = \sum_{j=0}^{q} \beta_j L^j$.

tools of frequency domain analysis which is harder to execute than time domain method does (Stock and Watson, 2011).

The second assumption is associated with the selection of form of the factor forecast model. Since the emphasis of this thesis is the h-month-ahead forecasts, Stock and Watson (2002b) suggested two possible approaches. The first approach is to develop a vector time series model for F_t and y_t and roll the (y_t, F_t) model forward. This approach entails a great number of parameters and in doing so the forecast performances would be eroded. The second approach is to recognise that a linear relationship exists between F_t and y_t (and its lags) and to construct the forecasts directly. This approach is better because it entails less parameters and and allows the estimation of factors to be done by PCA. Therefore, the econometric framework of factor models involves two steps: (1) the high dimensional dataset (the case of this thesis the 36 China's predictors) is linearly related to these unobservable and latent static factors in DI and FARIMA and dynamic factors in FAVAR; (2) China's macroeconomic variables of interests are also linearly related to these unobservable factors (both static and dynamic) and previous variables.

3.2.1 Assumptions underlining the diffusion index and theory of stationary data

The classical factor analysis has been widely used in psychology and social science but less so in disciplines of economics. This is perhaps because the assumptions made to classical factor analysis are too strong for economic data. The classical factor analysis assumes the factors and errors are serially and cross-sectionally uncorrelated[3], but economic data show a clear pattern of serial and cross-sectional correlation. Works on relaxing the assumptions of the classical factor model were

[3]The classical factor model has three assumptions: (1) e_t is iid over t; (2) N is fixed as T tends to be infinity; and (3) both F_t and e_t are normally distributed

conducted by Stock and Waston (1998, 2002a and 2005). Their efforts were to advance the theory of the static factor model towards to large approximate factor model. By large dimensional it means the N and T tend to be infinite regardless the ratio of N and T. Approximate means the factors and idiosyncratic errors are allowed to be serially and cross-sectionally correlated (weakly) which is more compatible to macroeconomic data.

To apply the factor model to China's data, the assumptions underlying the properties of stationary data are very important. Let F_t^0 and λ_t^0 denote the true factors and the loadings respectively and M be a generic constant. Consider the following five assumptions (Bai and Ng 2008)

Assumption F(0): $E \parallel F_t^0 \parallel^4 \leq M$ and $\frac{1}{T}\sum_{t=1}^{T} F_t^0 F_t^{0\prime} \rightarrow \sum_F > 0$ for an $r \times r$ non-random matrix $\sum_F$.

Assumption L: λ_i^0 is either deterministic such that $\parallel \lambda_i^0 \parallel \leq M$ or it is stochastic such that $E \parallel \lambda_i^0 \parallel^4 \leq M$. In either case, $N^{-1}\Lambda^{0\prime}\Lambda^0 \rightarrow \sum_\Lambda > 0$ for an $r \times r$ non-random matrix $\sum_F$ as $N \rightarrow \infty$.

Both assumptions $F(0)$ and L are moment conditions on the factors and the loadings and are standard assumptions in the factor models. Assumption $F(0)$ refers to the stationary factors and Assumption L concerns the factor loading. They jointly ensure that the factors are non-degenerate and each factor has a nontrivial contribution to the variance of X_{it}.

Assumption E:

(a) $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$

(b) $E(e_{it}e_{js}) = \sigma_{is,ts}$, $|\sigma_{is,ts}| \leq \overline{\sigma}_{ij}$ for all (t,s) and $|\sigma_{is,ts}| \leq \tau$ for all (i,j) such that $N^{-1}\sum_{i,j=1}^{N} \overline{\sigma}_{i,j} \leq M$, $T^{-1}\sum_{t,s=1}^{T} \overline{\tau}_{t,s} \leq M$ and $NT^{-1}\sum_{i,j,t,s=1} |\sigma_{i,j,t,s}| \leq M$

(c) For every (t,s), $E|N^{(-1/2)}\sum_{i=1}^{N}[e_{is}e_{it} - E(e_{is}e_{it})]|^4 \leq M$

(d) For each t, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \lambda_i e_{it} \rightarrow (0, \Gamma_t)$ as $N \rightarrow \infty$ where

$$\Gamma = \lim_{N\to\infty} \frac{1}{N} \sum_{s=1}^{T}\sum_{t=1}^{T} E(\lambda_t^{0\prime}\lambda_s^{0\prime} e_{is} e_{it}) \tag{3.5}$$

(e) For each i, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_t e_{it}(0,\Phi_i)$ as $T\to\infty$ where

$$\Phi_i = \lim_{T\to\infty} T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} E(F_t^0 F_s^{0\prime} e_{is} e_{it}) \tag{3.6}$$

Assumption E deals with the idiosyncratic errors. Part (a) assumes that the idiosyncratic errors is mean of zero and parts (b) and (c) assume the idiosyncratic can be weakly autocorrelated, weakly cross-sectional correlated, and heteroscedastic. Also, under the stationary data environment with $E(e_{it}e_{jt}) = \sigma_{ij}$, the eigenvalue of Ω is bounded by maximum value of $\sum_{j=1}^{N} = |\sigma_{i,j}|$ (Bai and Ng 2002).By assuming that the maximum eigenvalue of population covariance Ω is bounded and less than the generic constant M ($\sum_{j=1}^{N} = |\sigma_{i,j}| \leq M$). Therefore the proprieties of the large dimensonal approximate model are satisfied which means the macroeconomic data is compatible to the DI forecasts model (Bai and Ng, 2008). Parts (d) and (e) permit weak correlation between the factors and the idiosyncratic errors and also weak correlation between the loadings and the idiosyncratic errors.

Assumption LFE: (λ_i), (F_t) and (e_{it}) are three mutually independent groups.Dependence within each group is allowed.

Assumption IE: For all $t \leq$, $t \leq N$, $\sum_{s=1}^{T}|\Gamma_{s,t}| \leq M$, and $\sum_{i=1}^{T}|\overline{\sigma}_{ij}| \leq M$.

Assumption LFE means there exist three within-group dependences. They are (1) the factors F_t can be serially correlated (but cross-sectional independent), (2) the loadings $\lambda_i$can be serially correlated over time and (3) the idiosyncratic errors e_{it} can be both serially correlated and cross-sectional correlated. Note that there is no assumption imposed on the dependence between the factors and the loadings. Assumption IE strengthens assumption E in the sense that e_{it} is independent over time and cross-section independence, the assumption IE is implied by assumption E.

3.2.2 Estimating and selecting the factors

Two vitally important questions in applying the factor model to forecast variables of interest are: (1) what method is appropriate to estimate factors? and (2) what are the number of factors in subsequent regression?

Following Stock and Watson's (2002b) DI econometric framework, the estimation of the factor loading and factor scores are through the PCA. The PCA can deal with large or even infinite N and T regardless the ratio of N and T, and Stock and Watson (1998) suggested that the PCA is consistent estimation of the factors and the loadings even there is small amounts of data contamination as long as N is large. This suggestion makes PCA more compatible for China's data due to the weak quality of China's official data Fernald et al. (2014). Moreover, Stock and Watson (2011) and Bai and Ng (2008) show that if N is sufficiently large, then the factors estimated by PCA are precisely adequately to be the proxy of data in the subsequent forecasting model.

PCA is a data reduction technique which is used to re-express multivariate and large dimensional data with fewer dimensional components which contains the maximum amount of information from the original data. The method of PCA produces a $T\times k$ matrix of estimated factors[4] and a corresponding $N \times k$ of estimated factor loadings by solving the optimization problem $_{\Lambda^k, F^k} min$S(k)$where S(k) = (NT)^{-1} \sum_{i=1}^{N} \sum_{i=1}^{T} (x_{it} - \lambda_i^{k'} F_t^k)^2$(3.7)subject to the normalization that $\Lambda^{k'} \Lambda^k / N = I_k$ and $F^{k'} F^k$ being diagonal.

The only observable quantities is the dataset X_{it}, neither the factor scores, the factor loading, nor the idiosyncratic errors are physically observed. Factor scores and the factor loading in static factor model can be estimated simultaneously, that is, one can treat both factor scores and the factor loading as parameters (Bai and

[4]The number of factors estimated by principal component k does not necessarily equal to the true number of factor r

Ng, 2008). Let Σ_x be the covariance matrix of X_{it}, (λ_1, e_1),........,(λ_k, e_k) be the corresponding eigenvalue-eigenvector pairs of Σ_x where $\lambda_1 \geq \lambda_2 \geq$ .... $\geq \lambda_k$ and $e_i = (e_{i1}, ...e_{ik})'$. The i^{th} principal component of X_{it} is generated as follows:

Results: Let f_i denoted the i^{th} principal component of X_{it} for $i = 1, 2, ...k.$, we have

$$f_i = e_i' X_{it} \tag{3.8}$$

$$var(f_i) = e_i' \Sigma_x e_i = \lambda_i \tag{3.9}$$

$$Cov(f_i, f_j) = e_i' \Sigma_x e_i = 0 \tag{3.10}$$

Similarly, the estimation of the matrix of factor loadings is through

$$\Lambda \equiv [\Lambda_{ij}] = [\sqrt{\lambda_1 e_1} | \sqrt{\lambda_2 e_2} \cdots | \sqrt{\lambda_k e_k} \tag{3.11}$$

By equation (3.9) and (3.11), we have

$$\sum_{i=1}^{k} Var(X_{it}) = \sum_{i=1}^{k} \lambda_i = \sum_{i=1}^{k} Var(f_i) \tag{3.12}$$

The equation (3.12) implies that the proportion of total variance in X_{it} explained by the i^{th} principal component is simply the ratio between the i^{th} eigenvalue and the sum of all eigenvalues of covariance matrix of X_{it}.

With regards to determining the number of factors to retain in the subsequent factor forecasting models, Bai and Ng (2002) proposed three information criteria to determine the number of factors in the static factor model. Three information criteria trade off the benefit of including additional factors (or specifically additional parameters) in the candidate model against the cost of increased forecasts inaccuracy arising from additional parameters. So the Bai and Ng (2002)'s information criteria select the number of factors by minimizing a penalized likelihood function, where the penalty factor increases linearly with the number of factors. In theory, the Bai and Ng (2002)'s information criteria is based on the observation that the eigenvalue-eigenvector analysis of covariance matrix of China's macroeconomic dataset X_{it} and estimate the number of factors k that satisfies following optimisation problem

$$V_{(k)} = \min_{\Lambda, F^k} \frac{1}{NT} \sum_{i=1}^{N} \sum_{i=1}^{T} X_{it} - \lambda_i^k F_t^k)^2 \tag{3.13}$$

where k is the number of factors estimated by Bai and Ng (2002)'s information criteria, λ_i^k is k vector of factor loadings and F_t^k is k number of factor scores.

According to Bai and Ng (2002), if assumption F(0), L,E and LFE hold [5], the equation (3.13) leads to three information criteria:

$$IC_{p1}(k) = ln(V(k, \widehat{F}^k)) + k(\frac{N+T}{NT})ln(\frac{NT}{N+T}) \tag{3.14}$$

$$IC_{p2}(k) = ln(V(k, \widehat{F}^k)) + k(\frac{N+T}{NT})lnC_{NT}^2 \tag{3.15}$$

$$IC_{p3}(k) = ln(V(k, \widehat{F}^k)) + k(\frac{lnC_{NT}^2}{C_{NT}^2}) \tag{3.16}$$

There are some other methods to determine the number of factors such as scree plots which plots the ordered eigenvalues of $\widehat{\sum_x}$ against the rank of that eigenvalues. Scree plot is a useful method to access the marginal contribution of the i^{th} principal component to total variance of X_{it}; however, this method fails to measure the correlation between Y_t and X_{it}. In the other word, even through the first i^{th} component captures vast majority of variance from X_{it}, the correlation between Y_t and i number of component might be weak, resulting in eroding the appropriateness of subsequent forecasting models. On the other hand, Bai and Ng (2008) showed that if assumption F(0), L, E and LFE hold and the method of PCA is used to

[5] These assumptions state that in additional to standard assumptions on factor model, each factor has a nontrivial contribution to variance of X_{it}, allow for the limited time-series and cross-section dependence in the idiosyncratic component. Heteroskedasticity in both the time and cross-section dimensions is also allowed

estimated factor loadings and scores, their information criteria estimates k that is consistent with true number of factors r assuming that true value of r is finite and does not increase with N and T. Therefore, in what follows this study applies equation (3.14) to select the number of factors to retain in subsequent DI forecasting methodology.

3.3 The factor forecasting methodology

3.3.1 Diffusion index

The first statistical factor forecasting model in this study is the DI forecasting model developed by Stock and Watson (2002b). The distinctive feature of it is to add factors estimated by PCA to an otherwise AR model. It has the following form:

$$y_{t+h} = \alpha' \widetilde{F}_t + \beta' W_t + \epsilon_{t+h}, \tag{3.17}$$

where $h > 0$, $\beta' W_t$ is predetermined variables (the lags of dependent variables as of in the Stock and Watson (2002), the $\widetilde{F}_t$ is the factors estimated by PCA and the ϵ_{t+h} is the error.

Furthermore, Bai and Ng(2008) added the assumption FAR as stated below to the DI forecast model.

Assumption FAR:

(a) $Let z_t = (F_t' W_t')', E||z_t||^4 \leq M$; $E(\epsilon_{t+h}|y_t, z_t, y_{t-1}, z_{t-1}, ...) = 0$ for any $h > 0$; z_t and ϵ_t are independent of the idiosyncratic errors e_{is} for all i and s. Furthermore, $T^{-1} \sum_{t=1}^{T} z_t z_t' \xrightarrow{p} \sum_{zz} > 0$

(b) $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_t \epsilon_{t+h} \xrightarrow{d} N(0, \sum_{zz,\epsilon})$, where $\sum_{zz,\epsilon} = lim \frac{1}{T} \sum_{t=1}^{T} (\epsilon_{t+h}^2 z_t z_t') > 0$

With assumptions F(0), L, E, LFE, IE and FAR on hold, the forecasting equation

(7) is so-called DI forecasting methodology of Stock and Watson (2002b). Because DI is capable of exploiting the information in large dimensional panel data while keep the size of the forecasting model small, the DI forecasting methodology are indeed now used by various government agencies across a number of countries as well as for independent consultations and professional forecasts companies alike.

Following the framework of DI methodology of Stock and Watson (2002b), the one-month-ahead point forecasts for four China's variables are

$$\hat{y}_{t+1} = \hat{\alpha}_h + \beta_h \widehat{F}_t + \sum_{j=1}^{p} \hat{\gamma}_j y_{t-j+1}, \tag{3.18}$$

where $\widehat{F}_t$ is the vector of k factors estimated by PCA and selected by Bai and Ng (2002) information criteria.

The h-month-ahead point forecasts model is

$$\hat{y}_{t+h|t} = \hat{\alpha}_h + \beta_h \widehat{F}_t + \sum_{j=1}^{p} \hat{\gamma}_{hj} y_{t-j+1}, \tag{3.19}$$

where $\hat{y}_{t+h|t}$ is the h-step-ahead variable to be forecast, the constant term is introduced explicitly, $\widehat{F}_t$ are factors estimated by principal component, y_{t-j+1} is lags of variables, $\hat{\gamma}$ are coefficients associated with lags and the subscripts t reflect the time horizon of variable, and the lag of AR terms is selected by Akaike information criterion (AIC).

In order to examine the effectiveness of the statistical factor model further, two additional statistical factor models are conducted in this study. They are FARIMA and FAVAR models.

3.3.2 Factor-augmented autoregressive integrated moving average

FARIMA adds the factors esimated by PCA to the that of a standard ARIMA model. Similar to DI forecasting methodology, the FARMA also incorporates the latent factors from the China's economy and applies them to the subsequent forecasting model. However, the differences between FARMA and DI models is that FARMA incorporates a moving average term to explain China's leading macroeconomic variables.

The one-month-ahead factor-augmented ARIMA forecasting model has following form:

$$\hat{y}_{t+1} = \hat{\alpha}_h + \beta_h \widehat{F}_t + \sum_{j=1}^{p} \hat{\gamma}_j y_{t-j+1} + \sum_{i=1}^{q} \hat{\delta}_j \varepsilon_{t-i+1} + \varepsilon_t, \tag{3.20}$$

where $\widehat{F}_t$ is the vector of k factors estimated by PCA and selected by Bai and Ng (2002) information criteria, and the lag order of p in AR term and q in MA term are selected recursively by AIC with maximum lag order of 16.

3.3.3 Factor-augmented VAR

The VAR methodology is based on the specification of as many equations as there are variables in the system. Each variable is explained by its own past and the past of the other variables in the system. There are no exogenous variables other than deterministic variables. The FAVAR is a VAR model that one of the variables are factors and other variables are China's leading macroeconomic variables to forecast. In general, the FAVAR representation can be writen as:

$$\begin{bmatrix} F_t \\ X_t \end{bmatrix} = \begin{bmatrix} \Phi(L) & 0 \\ \Lambda\Phi(L) & D(L) \end{bmatrix} \begin{bmatrix} F_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \delta_t^F \\ \varepsilon_t^X \end{bmatrix}$$

where
$$\begin{bmatrix} \delta_t^F \\ \varepsilon_t^X \end{bmatrix} = \begin{bmatrix} I \\ \Lambda \end{bmatrix} \eta_t + \begin{bmatrix} 0 \\ \nu_t \end{bmatrix}$$

The one-month-ahead point forecast FAVAR is given by:

$$y_{t+1} = \sum_{k=0}^{p} a_{11} y_{t-k} + \sum_{k=0}^{p} a_{12} \widetilde{F}_{t-k} + \epsilon_{1t+1} \tag{3.21}$$

$$\widetilde{F}_{t+1} = \sum_{k=0}^{p} a_{21} y_{t-k} + \sum_{k=0}^{p} a_{22} \widetilde{F}_{t-k} + \epsilon_{2t+1} \tag{3.22}$$

and the h-month-ahead point forecast is given by:

$$y_{t+h} = \sum_{k=0}^{p} a_{11} y_{t-k+(h-1)} + \sum_{k=0}^{p} a_{12} \widetilde{F}_{t-k+(h+1)} + \epsilon_{1t+h} \tag{3.23}$$

$$\widetilde{F}_{t+h} = \sum_{k=0}^{p} a_{21} y_{t-k+(h-1)} + \sum_{k=0}^{p} a_{22} \widetilde{F}_{t-k+(h-1)} + \epsilon_{2t+h}, \tag{3.24}$$

where the lag order of FAVAR is determined by AIC and number of factor in $\widetilde{F}$ is estimated by PCA and determined by Bai and Ng (2002) information criteria.

3.4 Theoretical underpinnings of competing models

This study proposes three conventional time series models as competing model. The first competing model is an AR model with lag length determined by AIC up to a maximum of 24 lag. The second competing model is an ARMA model with order of p (lag in AR term) and q (lag in MA term) are selected by AIC up to a maximum of 12 lag respectively. The third competing model is a simple exponential smoothing model. Among these three univariate time series models, the AR model is widely used as benchmark model in forecasting journals and this study will use it as the benchmark model.

3.4.1 Autoregressive model

A natural starting point for a forecasting model is to use only the previous value of Y (that is Y_{t-1}, Y_{t-2}.) to forecast Y_t. This approach is known as AR forecasting model which specifics that the dependent variable depends linearly on its previous values. The AR model is defined as

$$Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i} + u_t, \tag{3.25}$$

where $E(u_t|Y_{t-1}, Y_{t-2}, ...) = 0$.

The lag length of AR forecasting model is selected recursively by AIC. Choosing the lag order in the AR model is a very influential decision as the forecasting performances is very sensitive to the lag order. On the one hand, if the lag order is too low, we will omit potentially valuable information contained in the more distant lagged values. On the other hand, if the lag order is too high, we will estimate more parameters than necessary and introduce additional forecasting errors. In theory, choosing the lag order p of an AR model requires balancing the marginal benefit of including more lags against the marginal cost of additional estimation models error. A practical and common way to determine the order of the lag in AR model is by minimizing an information criteria. One of the extensively used information criterion in forecasting literature is the Akaike information criteria (AIC). In practice, the model with minimum AIC is selected and used for forecasting. The one-month-ahead AR forecasting model is defined as:

$$\hat{y}^1_{T+1} = \hat{\alpha}_1 + \sum_{j=1}^{p} \hat{\gamma}_j y_{T-j+1} \tag{3.26}$$

where $\hat{\alpha_1}$ is the intercept term, the $\hat{\gamma}$ are the coefficients corresponding to the lags

y_{T-j+1} and n-month-ahead AR forecasting model is defined as:

$$\hat{y}^n_{T+n} = \hat{\alpha}_n + \sum_{j=1}^{p} \hat{\gamma}_j y_{T-j+n} \tag{3.27}$$

where $\hat{\alpha}_n$ is the intercept term, the $\hat{\gamma}$ are the coefficients corresponding to the lags y_{T-j+1}. Lag length are recursively selected by AIC with maximum lag order of 24.

3.4.2 Autoregressive integrated moving average model

Extending the AR model to ARMA model by considering the autocorrelation in the error term is straightforward. The ARMA model specifies that the dependent variable depends linearly on its previous values and some unobservable random error. The one-month-ahead forecast ARMA model is given by:

$$\hat{y}_{t+1} = \hat{\alpha}_h + \sum_{j=1}^{p} \hat{\gamma}_j y_{t-j+1} + \sum_{i=1}^{q} \hat{\delta}_j \varepsilon_{t-i+1} + \varepsilon_{t+1} \tag{3.28}$$

and h-month-ahead forecast is:

$$\hat{y}_{t+h} = \hat{\alpha}_h + \sum_{j=1}^{p} \hat{\gamma}_j y_{t-j+h} + \sum_{i=1}^{q} \hat{\delta}_j \varepsilon_{t-i+h} + \varepsilon_{t+h}, \tag{3.29}$$

where lag length of p and q are selected by AIC.

3.4.3 Exponential smoothing

Simple exponential smoothing model is also an univariate time series model that can predict variables of interest. The simple exponential smoothing model is a forecasting model where forecasts are calculated using weighted averages and where the weights decrease exponentially as observations come from further in the past - the smallest weights are associated with the oldest observations. The one-month-ahead exponential smoothing model is given by:

$$\widehat{y}_{t+1} = ay_t + a(1-a)y_{t-1} + a(1-a)^2 y_{t-2} \cdots \tag{3.30}$$

and the h-month-ahead forecast is given by:

$$\widehat{y}_{t+h} = ay_t + a(1-a)y_{t+h-1} + a(1-a)^2 y_{t+h-2} \cdots , \tag{3.31}$$

3.5 Forecasting performances evaluation

This section reviews the methods of forecasting performances evaluation related to AR forecasts, ARIMA forecasts, exponential smoothing forecasts, DI forecasts, FARMA forecasts and FAVAR forecasts. Such measures are indispensable for helping us to: (1)select the best model to forecast China's core macroeconomic variables among six candidate models, (2) evaluate and report the likely size of forecasting error.

This study focuses on the h-month-ahead prediction and the forecasting regressions are projections of an h-month-ahead variable y_{t+h} onto t-dated predictors. In order to compare the forecasting performances of six models, this study uses the most common forecasting performances' evaluation methods - the mean square forecast error (MSFE). The MSFE has been widely used to compare and examine the forecasting performances of forecasting models, see Stock and Watson (2002b), Giannone et al. (2008) and many others.

The MSFE gives the sum of the square of forecasts' error which allows analysis to compare the forecast performance in a easy-to-interpret-manner , regardless the sign of the forecasting errors in each month. In practice, the model with lowest MSFE is the best forecasting model.

The MSFE is defined as follows:

$$MSE = \frac{\sum(Y_t - F_t)^2}{n} = \frac{\sum e_t^2}{n}, \tag{3.32}$$

where Y_t is the actual data observed at period t, F_t is the forecast value generate by candidate models, n is the total number of out-sample-period, and e_t^2 is the difference between actual data and forecast value and is known as the forecast error at period t.

To express comparisons into a easy-to-interpret-manner, this study sets the MSFE of AR forecast as a benchmark and assigns the value of it as 1. Therefore, the range of the MSFEs are narrowed down to a small percentage number. However, the biggest disadvantage of the MSFE measure is because the MSFE is quadratic loss function; hence it is influenced dramatically by extreme values. To deal with this disadvantage, the proper data transformation and screening for the outlier are vitally important.

3.6 Conclusion

The intention of this chapter has been focused on reviewing the main theoretical underpinnings of the AR model and the factor model. The theoretical reviews of univariate time series models include estimation of parameters, lag length selection and form of AR, ARMA and exponential smoothing model for one-month-ahead and h-month-ahead forecasts. The theoretical reviews on the statistical factor model include the premise of factor model, the theory of stationary data, the estimation method through PCA, how determine the number of factors in static factor model though Bai and Ng (2002)'s information criteria and form of three factor forecasting models, factor-augmented ARMA and FAVAR one-month-ahead and h-month-ahead forecasts.

The well defined AR forecast model is regarded as the most common benchmark forecast in forecasting literature. It is simple model but empirically proved to be a useful model to forecast macroeconomic variables in many countries. The factor model, on the other hand, is one of a recent developed statistical model that can organise large dimensional data into a manageable manner. The premise of the factor model is based on the belief that a few latent and unobservable factors exist which jointly drive the co-movment of economies, and perhaps that there are only a few important factors that influence the variability of macroeconomic.

Following (Stock and Watson, 2002b), the estimation of the factor model is through two steps. In the first step, the latent and unobservable factor are estimated by PCA and selected by Bai and Ng (2002)' information criteria. In the second step, we incorporate these estimated factors to subsequent forecasting models. Then, the relative MSFE and is computed to examine the forecasting performances of six models.

Chapter 4

Data

4.1 Introduction

This chapter presents a detailed discussion of data, including the sources of data, technique of data transformation, and analysis of reasons for transformation. The reason for writing this chapter individually, in stead of classifying it as a section in results and discussion chapter, is because there is a widely held view academic literature that scepticism exists about the quality of China's data, which is worth discussing in a separate chapter. Literature on discussing quality of China's data can be seen in Rawski and Mead (1998), Chow (2006), Holz (2004), Holz and Lin (2001a), Holz (2008), Holz and Lin (2001b),and Nakamura and Steinsson (2014).

The scepticism about the unreliability of China's official data comes from the view that data falsification is conducted by National Bureau Statistics of China in order to meet economic growth targets and is increasingly the norm (Holz, 2004). Although there is no certain evidence on data falsification, it does not mean China's statistical system is necessarily honest in its statistical reporting. Also, even if National Bureau Statistics of China is not engaged to purposefully falsify the data, the margin of error in much of published data is likely to be sufficiently large to allow statistical authority to have a choice of final value to be reported from a relatively wide range of collected value (Holz, 2004).

Given the unreliability of China's official output data, this study intends to forecast economic activity indicators. Due to lack of availability of monthly GDP data[1], the forecasting literature usually forecasts industrial production instead. Although China's industrial production has also received a critique on reliability, Holz and Lin (2001a), this study still forecasts it because it is the most common output indicators in forecasting literature. In additional to forecast industrial production this study utilises two macroeconomic variables as a proxy of economic activity and inflation: production of electricity and producer price index:industrial goods.

Using the production of electricity as an indicator of economic activity is a sensible choice and has two benefits. Firstly, the production of electricity is not on the government target list; that is, neither does the statistical authority needs to self-motivatedly involve itself in falsifying the data in order to meet certain growth target nor does the statistical authority need to choose the margin error and final value of production of electricity that satisfies the media and public. Secondly, the information on the volume of production of electricity is relatively easier to collect than GDP or industrial production . To collect a GDP data, National Bureau Statistics of China employs professional data collectors to do surveys for every industry which may generate considerable error. Collecting the volume of production of electricity, by comparison, is a straightforward task - all data collectors need to do is go to power plant and read the volume number on the machines. In this way the size of collection errors is minimised. For these two reasons, production of electricity is an appropriate indicator of industrial production because high level of industrial production entails consuming large volume of electricity.

[1] GDP is available at quarterly and anually

Producer price index (PPI):industrial goods is indicator for the output and inflation. PPI: industrial goods measures the average changes in prices received by domestic producers for their output (industrial goods). The PPI:industrial is not a prefect substitute for the industrial production or rate of inflation, rather it provides information on how producers would price their industrial goods given the economic conditions. Normally, the producer would charge a high price when demands for industrial goods are high or rate of inflation is high. So PPI:industrial goods is an indicator that jointly reflects the economic activity and rate of inflation from the producer perspective.

4.2 Data description

This study uses a time series dataset covering monthly inflation rate measured as CPI, monthly volume of industrial production, the monthly volume of production of electricity and the producer price index:industrial and other 32 important macroeconomic variables. The complete dataset spans 1997:01 to 2014:03. The full dataset therefore contains 36 macroeconomic variables over the period between 1997:01 to 2014:03. All 36 macroeconomic variables can be found and downloaded at CEIC data manager[2]. Table 4.1 lists the names of the variables, their sources and transformation technique used to every series.

[2]CEIC is Macroeconomic, Industry, and Financial time series databases for Global Emerging and Developed Markets. It collects China's data from the National Bureau of Statistics of China

Table 4.1: Sources of data

	Variables	Sources
1	Government Revenue	CEIC
2	Government Expenditure	CEIC
3	Industrial Sales: Light Industry	CEIC
4	Industrial Sales: Heavy Industry	CEIC
5	Industrial Sales: State Owned	CEIC
6	Industrial Sales: Collective Ownership	CEIC
7	Consumer Price Index: MoM	CEIC
8	Production of Primary Energy: Crude Oil	CEIC
9	Production of Primary Energy: Natural Gas	CEIC
10	Production of Primary Energy: Electricity	CEIC
11	Financial Institution Deposits: Savings Deposits	CEIC
12	Index: Shanghai Stock Exchange: Composite	CEIC
13	Index: Shenzhen Stock Exchange: Composite	CEIC
14	Spot Exchange Rate: Period Avg: SAFE: RMB to US Dollar	CEIC
15	Spot Exchange Rate: Period Avg: SAFE: RMB to Japanese Yen	CEIC
16	Spot Exchange Rate: Period Avg: SAFE: RMB to Bistish Pound	CEIC
17	Money Supply M0	CEIC
18	Money Supply M1	CEIC
19	Money Supply M2	CEIC
20	Purchasing PI: Raw Materials (RM): Total	CEIC
21	Producer Price Index: Industrial Products	CEIC
22	PPI: IP: Producer Goods	CEIC
23	Retail Price Index	CEIC
24	Retail Price Index:Urban	CEIC
25	Retail Price Index: Rural	CEIC
26	Industrial production	CEIC
27	Interest rate: discount rate	CEIC
28	Interest rate:lending rate	CEIC
29	Interest rate: borrowing rate	CEIC
30	Oil price	CEIC
31	Commodity Agricultural Raw Materials Index	CEIC
32	Commodity Metals Price Index Monthly Price	CEIC
33	CN: Export FOB	CEIC
34	CN: Import CIF	CEIC
35	CN: Effective Exchange Rate Index: BIS: Real	CEIC
36	CN: Effective Exchange Rate Index: BIS: Nominal	CEIC

Notes:CEIC is Macroeconomic, Industry, and Financial time series databases for Global Emerging and Developed Markets. It collects China's data from NBSC the National Bureau of Statistics of China.

Data on the CPI is obtained in CEIC data manager and is month-on-month basis. That is, the monthly CPI is computed by setting up the previous month CPI to 100 and thus each month's CPI is a proxy of the actual rate of inflation during one month.

Data on the monthly industrial production and monthly production of electricity is the actual volume of productions collected by the National Bureau Statistics of China. Data on producer price index:industrial product is a index number which sets up previous year value equal to 100. Data for other 32 macroeconomic variables include measures of financial policy such as government revenue and government expenditure, real activity such as industrial sales and production of crude oil, stock prices index, exchange rates, monetary policy such as money supply and interest rate, commodity such agriculture and metals price, volume of international trade, and world oil price.

4.3 Data transformation

The theory outlined in Chapter 3 states that stationarity of X_{it} is moment condition for factor analysis (Bai and Ng, 2008)[3]. So these 36 series were subjected to three preliminary steps: possible transformation by taking first difference, possible transformation by taking the first difference of logarithms implying the growth rate of the variables, and screening for the possible outliers[4]. The decision to take first difference or the difference of the logarithms was made judgmentally including inspection of the time series plot of the data and the unit root test. In general, the first differences of logarithms are taken to the series with the actual quantity that are not already in index or percentage, and first differences are taken to those series that are already in index or percentage number. The transformation details for each series are reported in Table 4.2.After these transformations, all series were further standardised to have a sample mean zero and a unit sample variance. These

[3]In theory, if any of 36 series is integrated or non-stationary, then the covariance matrix of X_{it} does not exit thus there is no a factor representation at all

[4]Screening for outliers is other vitally important step. Outliers can have deleterious effects on statistical analysis of regression results and performances of forecasting. First, they generally increase the variance of data and hence reduce the power of statistical tests such as t-test for coefficients. Secondly, they can seriously bias or influence the estimated forecasts value and forecasts error (Osborne and Overbay, 2004).

data transformations satisfy the assumption of stationartity of X_{it}. Finally, the transformed data were screened for outliers. Following Stock and Watson (2002b), observations that exceeded 10 times of the series mean were replaced by the series mean.

Table 4.2: Transformation for each series

	Variables	Transformation
1	Government Revenue	DL
2	Government Expenditure	DL
3	Industrial Sales: Light Industry	DL
4	Industrial Sales: Heavy Industry	DL
5	Industrial Sales: State Owned	DL
6	Industrial Sales: Collective Ownership	DL
7	Consumer Price Index: MoM	D
8	Production of Primary Energy: Crude Oil	DL
9	Production of Primary Energy: Natural Gas	DL
10	Production of Primary Energy: Electricity	DL
11	Financial Institution Deposits: Savings Deposits	DL
12	Index: Shanghai Stock Exchange: Composite	DL
13	Index: Shenzhen Stock Exchange: Composite	DL
14	Spot Exchange Rate: Period Avg: SAFE: RMB to US Dollar	DL
15	Spot Exchange Rate: Period Avg: SAFE: RMB to Japanese Yen	DL
16	Spot Exchange Rate: Period Avg: SAFE: RMB to Bistish Pound	DL
17	Money Supply M0	DL
18	Money Supply M1	DL
19	Money Supply M2	DL
20	Purchasing PI: Raw Materials (RM): Total	D
21	Producer Price Index: Industrial Products	D
22	PPI: IP: Producer Goods	D
23	Retail Price Index	D
24	Retail Price Index:Urban	D
25	Retail Price Index: Rural	D
26	Industrial production	DL
27	Interest rate: discount rate	D
28	Interest rate:lending rate	D
29	Interest rate: borrowing rate	D
30	Oil price	DL
31	Commodity Agricultural Raw Materials Index	D
32	Commodity Metals Price Index Monthly Price	D
33	CN: Export FOB	DL
34	CN: Import CIF	DL
35	CN: Effective Exchange Rate Index: BIS: Real	DL
36	CN: Effective Exchange Rate Index: BIS: Nominal	DL

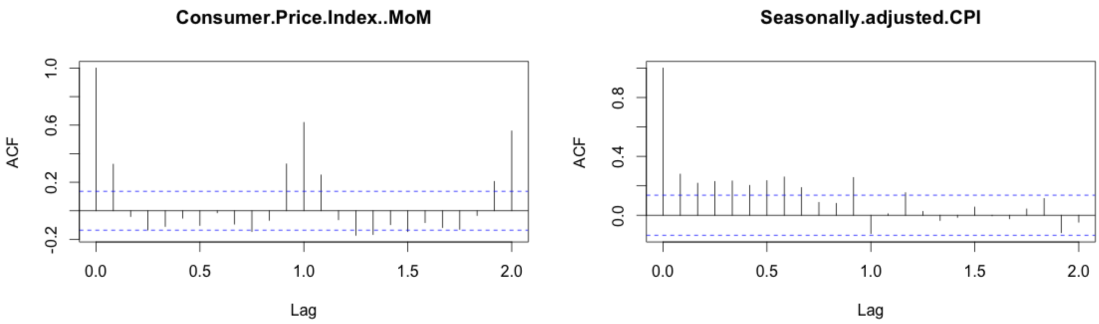
Note: D means the series is taken the first difference, and DL means the series is taken the first difference of logarithms

Another noteworthy point is the effect of Chinese New Year on the 4 series. The primary step taken to eliminate such effect in this thesis is through *R package X13-seasonal-R interface to X-13ARIMA-SEATS*. Although it is a ideal to include a dummy variable that reflect the effect of Chinese New Year effect, two main reasons make it unpractical. First, the X-13 package is sufficiently enough to soften such effect as proved by Figure 4.1, and 4.5. Second, the inclusion of dummy variable into model makes little contribution to solve the Chinese New Year effect but it increases the uncertainty of forecasts as the number of predictors in models goes to larger.

4.3.1 Consumer price index

In terms of the transformation of four real variables, the CPI to be predicted in this study is the seasonally adjusted month-on-month CPI. The reason to take seasonal adjustment to the CPI is because separating the seasonal components from the month-on-month CPI allows us to remove the effect of China's New Year. As China's New Year typically jumps between January and February, the seasonal adjustment is a sensible approach to remove the fluctuation of China's New Year and therefore clearly reveal the magnitude and direction of the CPI from month to month. The effect of seasonal adjustment is confirmed by Figure 4.1.

Figure 4.1: Autocorrelation function of adjusted and unadjusted CPI



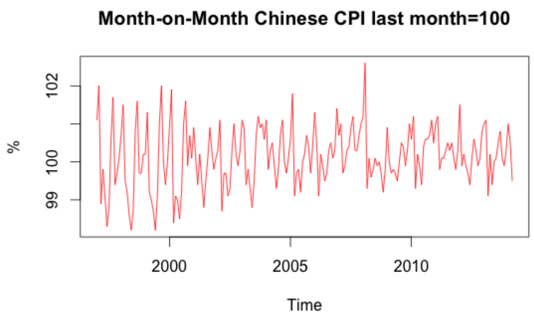
(a) ACF of unadjusted CPI (b) ACF of adjusted CPI

Notes: This seasonally adjusted CPI is transferred in R by the package: X13-seasaonl-R interface to X-13ARIMA-SEATS. The package performs a seasonal adjustment that works well in most circumstances.

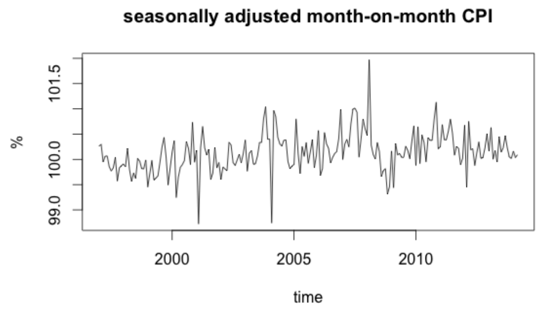
Figure 4.1 depict the autocorrelation function (A CF) of Month-on-Month CPI (in the left) and seasonally adjusted CPI (in the right) respectively. As can be seen the ACF of month-on-month CPI (a in Figure 4.1) clearly reveals that current month-on-month CPI is weakly correlated with the past 10 months' CPI but is strongly correlated with last year CPI and even correlated with CPI in two years ago. In contrast, the ACF of seasonally adjusted month-on-month CPI (b in Figure 4.1) provides a better explanation of current CPI as it suggests that the current CPI is correlated with previous CPI over the past eight months and has nothing to do with last year or last two years' CPI. Also, Figure 4.2 shows time series plots of month-on-month CPI and seasonally adjusted CPI. The plot of seasonally adjusted CPI clearly shows a smooth version except period during 2007-2008[5].

[5]This big structural break might be due to the global financial crisis

Figure 4.2: Time series plots of unadjusted and adjusted CPI



(a) Unadjusted CPI



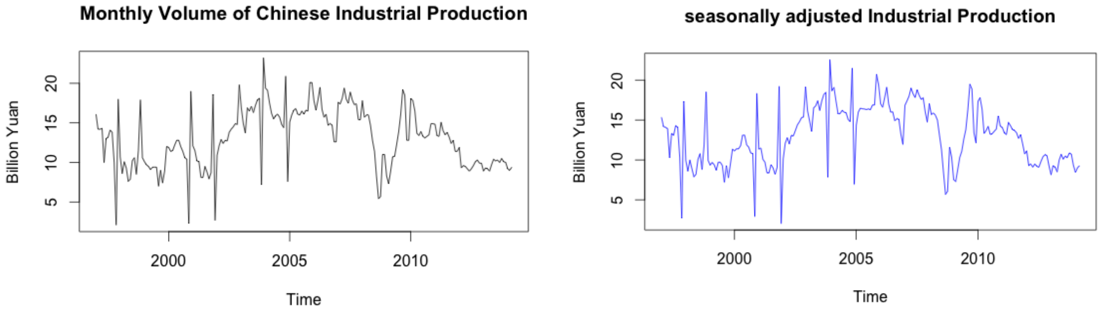
(b) Adjusted CPI

Therefore, the seasonal adjustment successfully removes the effect of China's New Year on CPI. In addition to seasonal adjustment, the CPI is modeled as being I(1). Although the seasonally adjusted CPI itself is stationary, but not every within-sample period is stationary. Thus modelling CPI as I(1) ensures the CPI in every with-in-sample periods are stationary.

4.3.2 Industrial production

Industrial production is a measure of the output of the industrial sector of the economy. The industrial sector includes manufacturing, mining, and utilities. Because industrial production is highly sensitive to interest rates and consumer demand, it often serves an important tool for forecasting GDP and economic activity. Unlike the CPI and production of electricity, the industrial production does not need to be seasonally adjusted. Figure 4.3 displays time series plots of unadjusted and adjusted monthly industrial production.

Figure 4.3: Time series plots of industrial production

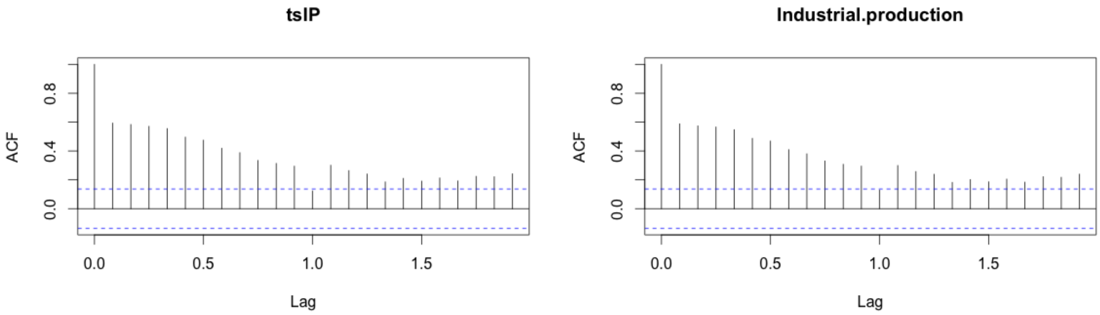


(a) Unadjusted industrial production (b) Adjusted industrial production

Notes: This seasonally adjusted industrial production is conducted by decompose() function in R. The X13-seasonal package returns exact same value before and after seasonal adjustment.

Figure 4.4 shows the ACF of unadjusted and adjusted industrial production.

Figure 4.4: ACF of industrial production



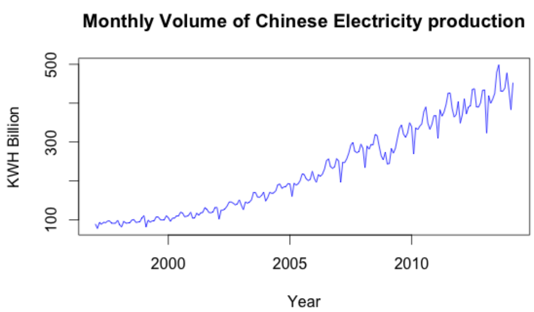
(a) Unadjusted Industrial Production (b) Adjusted Industrial Production

As shown in Figure 4.4, seasonal adjusted has no effects on the ACF of industrial production which means there is no need for the seasonal adjustment. To be consistent with Stock and Watson (2002b) and many other forecasting literature, the industrial production is transformed into the difference of the logarithm, implying the growth rate of industrial production.

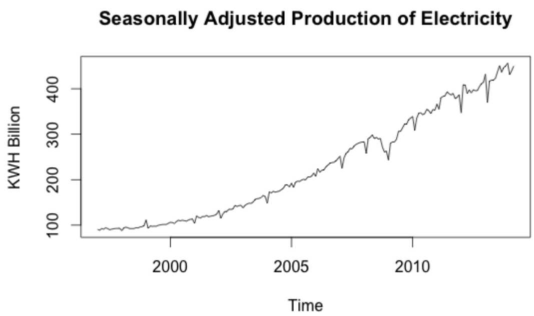
4.3.3 Production of electricity

Like the CPI, the monthly production of electricity needs to be seasonally adjusted. Figure 4.5 reveals time series plots of unadjusted and adjusted production of electricity. Looking at unadjusted plot at Figure 4.5 (a), there are a great number of spike during each year. These extreme fluctuations could be explained by the large amount of electricity demand in summer and the low demand in the winter[6].

Figure 4.5: Time series plots of unadjusted and adjusted production of electricity



(a) Unadjusted production of electricity



(b) Adjusted production of electricity

Note: This seasonally adjusted production of electricity is conducted by the R package X13-seasonal.

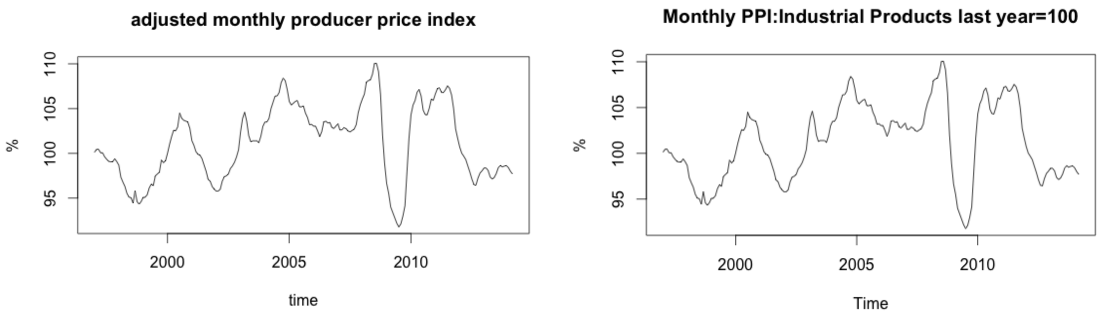
The seasonally adjusted monthly production of electricity in Figure 4.5 (b) ,on the other hand, represents the smoothed production of electricity which removes the seasonal effects. Furthermore, the seasonally adjusted production of electricity is modelled as being the first difference of logarithms, representing the growth rate of production of electricity each month. By doing so, the series are stationary in every with-in-sample period.

[6]From my experiences, households and businesses need to turn the air conditioning on in order to resist the hot summer weather and this applies to the vast majority of cities. However, in winter, households and business tend not to use heater to warm them up, especially in South part of China. Therefore, the demand and usage of electricity is high in summer and low in winter

4.3.4 Producer price index:industrial goods

The data transformation techniques applied to PPI:industrial goods[7] differ from those applied to CPI. The seasonal adjustment has no effects on the PPI:industrial goods. This is proved by Figure 4.6 and 4.7.

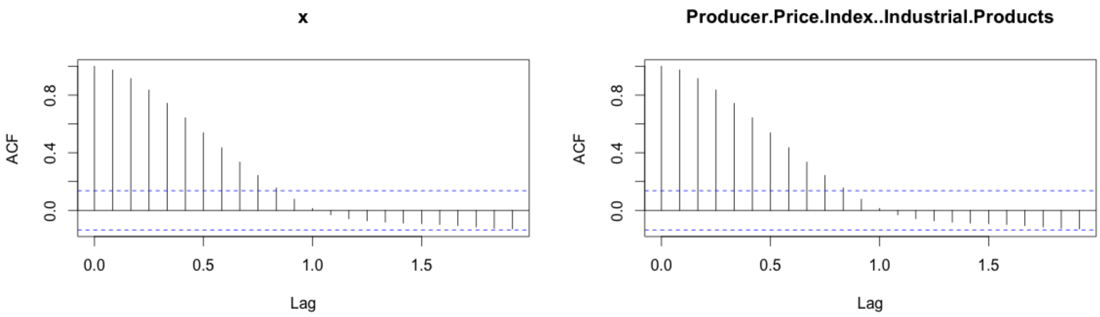
Figure 4.6: Time series plots of PPI:industrial goods



(a) Unadjusted PPI:industrial goods (b) Adjusted PPI:industrial goods

Notes: The X13-seasonal package and decompose() function in R have been used to detect the seasonal effects. None of them produce adjusted series.

Figure 4.7: ACF of PPI:industrial goods



(a) Unadjusted PPI:industrial goods (b) Adjusted PPI:industrial goods

Figure 4.8 shows the time series plots of adjusted and unadjusted PPI:industrial goods. Figure 4.9 depicts ACF of them. Both figures demonstrate that there is no need to take seasonal adjustment to PPI:industrial goods.

[7] The base period of the monthly PPI series reported in the National Bureau of Statistics of China (NBSC) is the same month of the previous year, which differs from the US standards. For example, the base period of PPI in January 2010 is January 2009 and the base period of CPI in February 2010 is February 2009.

4.4 Conclusion

This chapter discusses the sources of the data, and the techniques applied to transfer data including taking difference of logarithms or first difference, and applying seasonal adjustment to four leading China's macroeconomic variables. The motivation of using production of electricity and PPI:industrial goods as proxy of economic activity are discussed in depth.

Due to the extensive critique on the quality of China's officially reported output, this study forecasts three economic indicators that are believed to be appropriate substitutions for GDP. The first indicator is industrial production. Although there are still some criticism on quality of reported industrial production figures (Holz and Lin, 2001a), the level of motivation of statistical authorities to falsify the data is less than to falsify GDP, partly because there is no certain industrial production target that they have to meet and partly because media and public pay less interest to industrial production than to GDP. Production of electricity is a sensible indicators when the quality of industrial production is questioned. Strong economic growth normally requires a high level of electricity consumption. PPI:industrial goods is an appropriate indicator to reflect both economic activity and inflation from producer perspective.

Consistent with theories outlined in Chapter 3 and many forecasting literature such as Stock and Watson (2002b) and Forni et al. (2003), the X_{it} needs to be I(0) process. If it is not, then the covariance matrix of X_{it} does not exist, which means there is no factor representation at all. Hence, the transformation of data are processed as follows. In step one, possible transformation by taking difference of logarithms is taken to the variables that are not in percentage. In step two, possible transformation by difference is taken to the variables that are already in percentage or index. Step three is about screening for outliers. After these transformations, all series were further standardised to have a zero mean and unit variance.

As for four leading macroeconomic variables, the CPI firstly is being seasonally adjusted and is processed being I(0). The industrial production is taken difference of logarithms, implying the growth rate. Production of electricity is also seasonally adjusted and transferred by taking difference of logarithms. The PPI:industrial goods is transferred by differencing.

Chapter 5

Results and discussion

5.1 Introduction

This chapter provides the detailed empirical results of forecasting of four China's leading variables for six models. In order to access the forecasting performances of the six models, the results of relative MSFE are reported.

The emphasises of the discussion section are on the twofold: (1) the hypothesis testing, looking at whether the statistical factors forecasts can contribute significant improvements over AR forecasts[1]? and (2) a surprising but sensible results by the exponential smoothing forecasts and FAVAR forecasts.

This chapter is organized as follows: Section 5.2 describes the forecast experimental design including estimation of parameters and factors, the model selection and so forth. Section 5.3 presents the empirical results of forecasting measured by relative MSFE discuss the results in depth. Section 5.4 draw a concluding comment.

[1] By significant improvements I define it by the majority of factor forecasts should contribute at least 20% forecasting error reduction. For instance, in Stock and Watson (2002b) they found that for 12-month-ahead forecasting horizon the 15 out of 17 DI and DI forecasts outperform AR forecast at more than 30%

5.2 Forecast experimental design

The forecast comparison is performed in a simulated out-of-sample frame-work where all statistical calculations are done using a fully recursive methodology. To ensure sufficient within sample and out of sample forecasting period, the six models are first estimated on data from 1997:02 to 2005:08[2]. The h-step-ahead forecasts are then computed based on predictors and factors in 2005:08. The estimation sample period is then augmented by one month and the corresponding h-step-ahead forecast is computed again. This process repeats all the way up until the final out-of-sample forecast (2014:03) was computed based on predictors and factors on 2014:02 (for one-month-ahead), 2013:12 (for three-month-ahead) and 2013:09 (for six-month-ahead). Every month (i.e. for every augmentation of the sample period) all model estimations, standardization of the data, calculation of the estimated factors, etc., are repeated. Finally, out-of-sample MSFE is then computed in order to compare the forecasting performances of 6 candidate models.

5.3 Empirical results and discussion

5.3.1 One-month-ahead forecasts

Empirical results

The relative MSFEs for the four China's leading variables are reported for one-month-ahead forecasts in detail in Table 5.1[3].

[2]The full data spans from 1997:01 to 2014:03, the estimation starts with 1997:02 is because 4 series have been transformed either by difference of difference of logarithm

[3]where CPI denotes to consumer price index, IP means industrial production. PE refers to production of electricity. PPI is producer price index: industrial goods. AR is referred to autoregressive model, ARIMA means autoregressive integrated moving average model, ES denotes to exponential smoothing model, DI means difussion index model. FARIMA is factor-augmented autoregrssive integrated moving average model. FAVAR is factor-augmented vector autoregressive.

Table 5.1: Relative mean square forecasting errors for one-month-ahead forecast

	CPI	IP	PE	PPI
AR	1.00	1.00	1.00	1.00
ARIMA	0.89	1.09	0.95	0.74
ES	1.67	0.55	1.18	1.11
DI	1.15	1.15	0.98	0.87
FARIMA	0.92	1.09	1.23	0.82
FAVAR	1.26	0.80	0.96	0.89

The relative MSFE is computed relatively to MSFE of AR forecasts so that the AR forecasting model has a relative MSFE of 1.00. For example, the out-of-sample MSFE of the DI model of CPI is 115% that of AR forecast of CPI at the one-month-ahead forecasting horizon. According to Table 5.1, the ARIMA models generate the best forecasts except for the industrial production series. As can be seen, ARIMA forecasts of CPI and PPI series significantly outperform the benchmark AR model (0.89% and 0.75% of AR forecasts respectively) and ARIMA forecasts of production of electricity moderately outperform the benchmark AR model (0.95% of AR forecast). Exponential smoothing produces the worst forecasts except for industrial production series.

In terms of the forecasting performances of the statistical factor models, the first comparison is between the DI forecasts and the AR forecasts. DI forecasts of CPI and industrial production series apparently underperform the benchmark AR forecasts, producing 15% more forecasting errors. The DI forecast for production of the electricity series barely outperforms the benchmark AR model (2% better) and the DI forecast for PPI:industrial goods series dramatically outperforms the benchmark AR model (13% better).

The second comparison considers the difference between FARIMA forecasts and benchmark AR forecasts. Both CPI and PPI produce FARIMA forecasts that are better than those of AR forecasts (8% and 18% better respectively) while industrial production and production of electricity generates FARIMA forecasts that are worse than those of AR forecasts (9% and 23% worse).

The third comparison considers the differences between FAVAR model and benchmark AR model. It is apparent that FAVAR forecasts of IP, production of electricity and PPI:industrial goods series outperform the benchmark AR forecasts while FAVAR of CPI forecasts significantly underperform the benchmark AR model. In some case, the improvement of factor models over benchmark AR model are substantial; for example, for industrial production, the FAVAR forecast only produces 80% of MSFE than that of the benchmark AR model. FARIMA forecast of PPI is 82% of that of AR forecast.

Discussion

One-month-ahead results suggest that the performances of comparable models is indefinite when the estimated factors are used. On the one hand, DI forecasts of production of electricity and PPI series outperform the benchmark AR forecasts which are favourable for factor forecasts. On the other hand, the DI forecast of CPI and IP underperform substantially those of AR models at 15% more MSFE. However, the performances of forecasts is usually better when a dynamic vector structure exists between lag factors and lag variables.

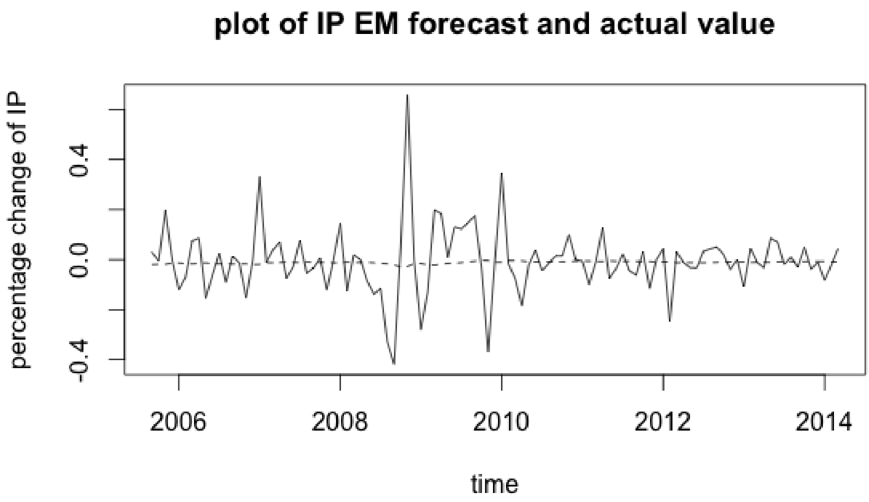
Inspection of Table 5.1 reveals a striking finding: simply using DI or FARIMA forecasts is not guaranteed to generate improvements over benchmark AR forecasts except for PPI series; rather, the forecasting performances are usually improved by incorporating estimated factors to a VAR model except for CPI series.

The performances of DI forecasts is inconsistent with Stock and Watson (2002b). They found the DI model to be very useful to improve forecasting accuracy for US macroeconomic variables but this is not the case in China. The performances of FAVAR forecasts, on the other hand, is consistent with Fernald et al. (2014) and

Bernanke et al. (2004).

Another notable result is that exponential smoothing generates the best forecasts for industrial production series - far more than other five models. The exponential smoothing basically generate a forecast that is almost same as naive forecast as shown in Figure 5.1.

Figure 5.1: Plot of exponential smoothing forecast versus actual for IP



The dotted line is the forecast and the solid line is the actual value. This applies to remaining figures in the Chapter.

As can be seen from Figure 5.1, the exponential smoothing model generates almost a mean zero forecast across out-of-sample period (a naive forecasting) which implies that the best forecasts of industrial production for the next month is simply today's industrial production.

5.3.2 Three-month-ahead forecasts

Empirical results

Table 5.2 reports the relative MSFEs three-month-ahead forecasts.

Table 5.2: Relative mean square forecasting errors for three-month-ahead forecast

	CPI	IP	PE	PPI
AR	1.00	1.00	1.00	1.00
ARIMA	0.72	1.48	1.07	0.85
ES	0.75	0.69	0.93	0.63
DI	1.00	1.17	1.05	0.66
FARIMA	0.75	1.50	1.05	0.61
FAVAR	0.69	0.74	0.83	0.77

According to Table 5.2, the competition between ARIMA forecasts and benchmark AR forecasts ends in a tie, as CPI and PPI forecasts significantly outperform the benchmark AR forecasts (28% and 15% better) while industrial production and production of electricity dramatically underperform compared to the AR forecasts (48% and 7% worse). In contrast to the performances of one-month-ahead forecasts, the exponential smoothing model dramatically outperforms the benchmark AR model for all four series. In the most case, the improvements are considerable. The MSFE of exponential smoothing of the CPI and the industrial production are only 75% and 69% of AR's respectively.

With regards to the forecasting performances of the statistical factor models, the DI model seems not improving forecasting accuracy over the benchmark AR model except for PPI:industrial goods. More precisely, the MSFE of the DI forecasts of the industrial production and the production of electricity are greater than those of AR forecasts' (17% and 5% more). The MSFE of the DI forecast of the CPI is same as AR forecast. The FARIMA model contributes a substantial improvement for CPI (25%) and PPI (39%) series while it produces extra forecasting error for industrial production (50%) and production of electricity (5%). Unlike performances of DI and FARIMA, FAVAR forecasts reveal significant forecasting improvements (at least 15%) over benchmark AR forecasts for all four China's leading macroeconomic variables.

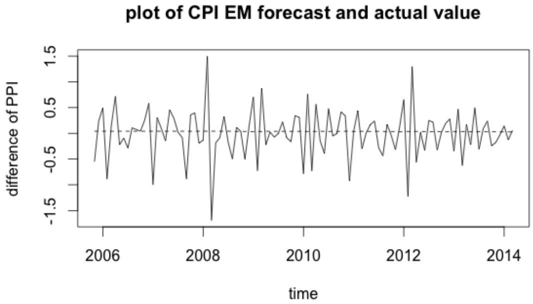
Discussion

The performances of DI and FARIMA model in three-month-ahead forecasting is vague. One the one hand, DI and the FARIMA show superiority over the benchmark AR model in some cases. For instance, DI forecast of PPI:industrial goods and FARIMA forecasts of CPI significantly outperform the AR forecasts at 34% and 25% respectively. On the other hand, DI and FARIMA produces some underperformed results such as FARIMA of industrial production (50% worse) and DI forecast of production of electricity (5% worse). FAVAR model, however, demonstrates superior forecasting performances than AR forecasts for all 4 series.

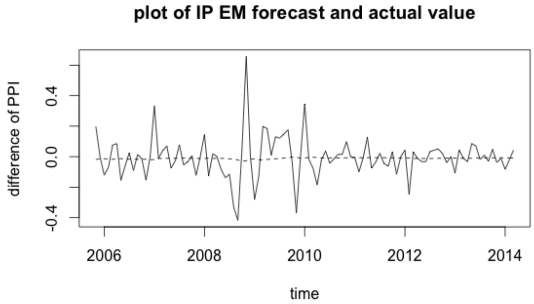
Inspection of Table 5.2 reveals an apparent finding: augmenting contemporaneous factors to standard AR and ARIMA models does not necessarily improve forecasting accuracy for CPI, industrial production and production of electricity. In some cases, it even produces worse results. Rather, the predictable dynamic of four series can be explained by vector structure of lag factors and lag series (the FAVAR model). Similar to results of one-month-ahead forecast, this finding is inconsistent with Stock and Watson (2002b) but is consistent with Fernald et al. (2014) and Bernanke et al. (2004). The reason for the superiority of FAVAR might be because of the weak quality of China's official reported data and rapid change of China's economic structure, which makes FAVAR very appropriate.

Another notable result is concerned with performances of exponential smoothing and FAVAR as shown in figure 5.2 and figure 5.3:

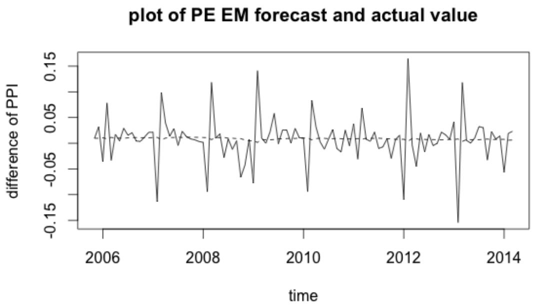
Figure 5.2: Plot of EM forecasts versus actual value



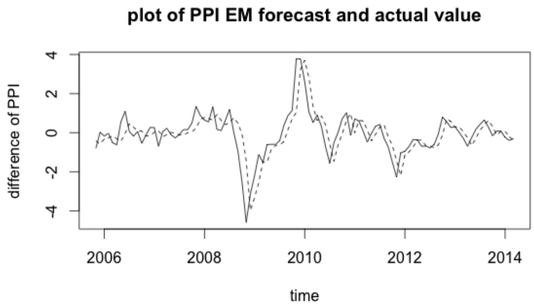
(a) CPI



(b) Industrial production



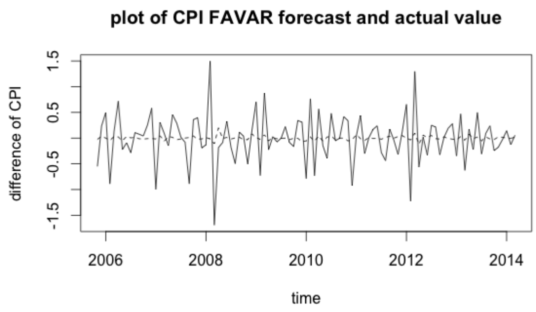
(c) Production of electricity



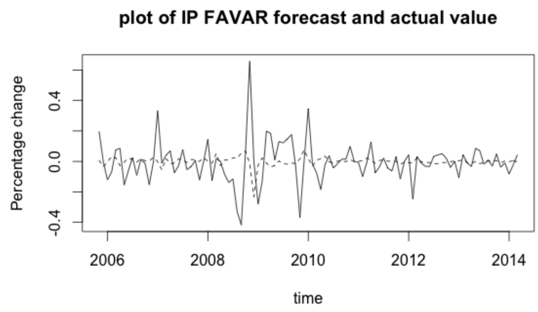
(d) PPI:industrial goods

As can be seen from Figure 5.2, the exponential smoothing model generates straight lines forecasts around zero except for the PPI series. This implies that the exponential smoothing forecasts are naive forecasts. As for PPI series, the exponential smoothing model generate a forecast that is one period lag to actual value.

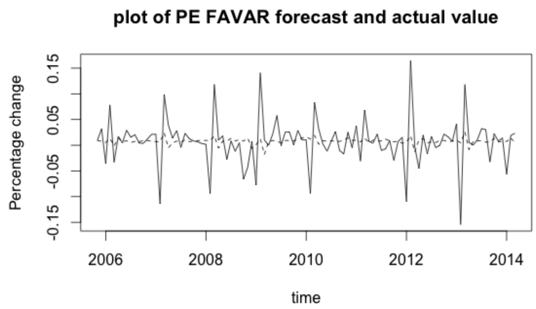
Figure 5.3: Plot of FAVAR forecast vs actual value



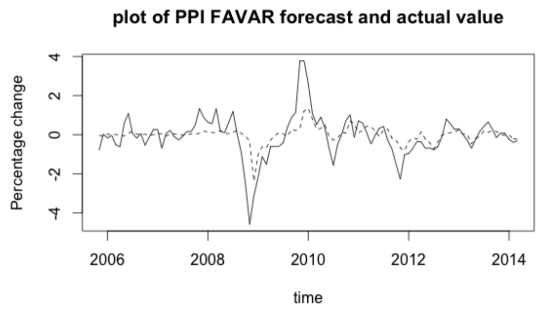
(a) CPI



(b) Industrial Production



(c) Production of Electricity



(d) PPI:industrial goods

The forecasting performances of the FAVAR model is somehow similar to those of forecasts of exponential smoothing models. According to Figure 5.3, the FAVAR forecasts of CPI and PE series generate naive forecasts implying that the predicted differences of rate of inflation and growth rates of production electricity are almost zero for entire out-of-sample forecast periods. The FAVAR forecast of industrial production series generates a naive forecast except for the period between late of 2008 and early of 2009.

Results from Figure 5.2 and 5.3 are surprising and unexpected because they essentially indicate that the better approach to reduce the forecasting errors and increase the predictable dynamic for the CPI, industrial production and production of electricity three-month-ahead is to look at today's CPI, industrial production and production of electricity.

5.3.3 Six-month-ahead forecasts

Empirical resutls

The relative MSFEs for four leading variables in six-month-ahead is given in Table 5.3

Table 5.3: Relative mean square forecasting errors for six-month-ahead forecast

	CPI	IP	PE	PPI
AR	1.00	1.00	1.00	1.00
ARIMA	0.89	1.30	1.18	0.97
ES	0.81	0.84	1.01	0.48
DI	1.02	1.02	1.04	0.54
FARIMA	0.91	1.27	1.09	0.52
FAVAR	0.81	0.82	0.99	0.81

According to Table 5.3, ARIMA forecasts of four series end a draw with benchmark AR forecast. More specifically, ARIMA forecasts of CPI and PPI outperform the benchmark AR forecasts at 11% and 3% respectively while industrial production (30% worse) and production of electricity (18%) significantly underperform to benchmark AR forecasts. Similar to performances in three-month-ahead, the exponential smoothing forecasts show marked improvements over benchmark AR forecasts, with one exception being production of electricity marginally worse (1%) than that of AR forecast.

With regards to the forecasting performances of the statistical factor models, both DI and FARIMA show superior and inferior performances. The only DI forecast that has better performance than AR forecasts is the PPI series (producing 54% MSFE of AR forecast) while DI forecasts of CPI, industrial production and production of electricity are slightly worse than those of AR forecasts. FARIMA forecasts of CPI (0.91% MSFE of AR forecast) and PPI (0.52% MSFE of AR forecast) have better forecasting performances than AR forecasts while industrial production (27% worse) and production of electricity (9%) generate worse forecasting performances than AR forecasts.

Unlike performances of DI and FARIMA forecasts, FAVAR forecasts of four variables show superior forecasting ability over benchmark AR forecasts. In some case, the improvements are substantial. The FAVAR of CPI and PPI, for instance, generate 81% forecasting error that of AR forecast, representing 19% improvement.

Discussion

According to results of relative MSFE in six-month-ahead forecasts, three out of four DI forecasts[4] underperform the benchmark AR forecasts, and two out four FARIMA forecasts[5] are better than those of AR forecasts. This implies that simply augmenting contemporaneous factors to the standard AR and ARIMA model does not necessarily improve forecasting accuracy for the CPI, the industrial production and the production of electricity series. In fact, augmenting contemporaneous factors to AR even generates worse performances. This results is inconsistent with Stock and Watson (2002b), suggesting DI and FARIMA forecasting methodology might not to be a sensible way to improve forecasting performances except for PPI:industrial goods series. The FAVAR forecasts, on the other hand, reveals substantial improvements over benchmark AR forecasts for all four series.

These results suggest the contemporaneous factors can not explain predictable dynamic of CPI, industrial production and production of electricity series. The performances of forecasts are always better when there exist a dynamic vector structure between lag factors and lag variables (FAVAR model). This is consistent with Bernanke et al. (2004) and Fernald et al. (2014).

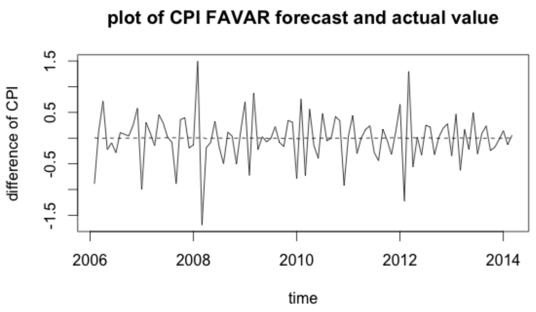
Another notable result is that, for the CPI, the industrial production and the production of electricity, the exponential smoothing and the FAVAR forecasts essentially

[4]they are the CPI, the industrial production and the electricity production

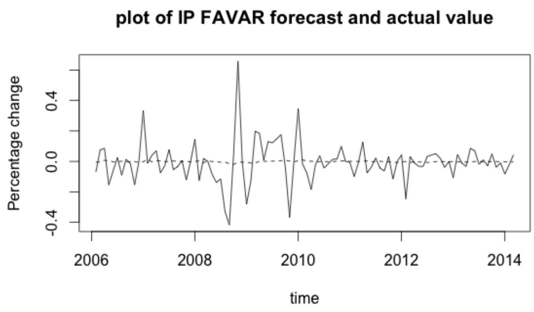
[5]the industrial production and production of electricity

generates naive forecasts as shown in Figure 5.4 below:

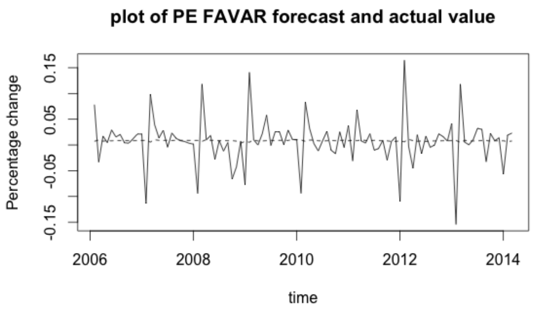
Figure 5.4: Plot of FAVAR and exponential smoothing forecasts versus actual value



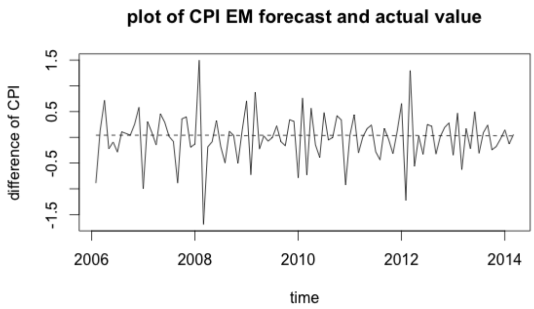
(a) CPI



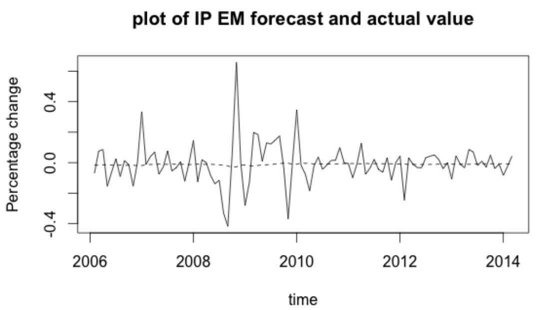
(b) Industrial production



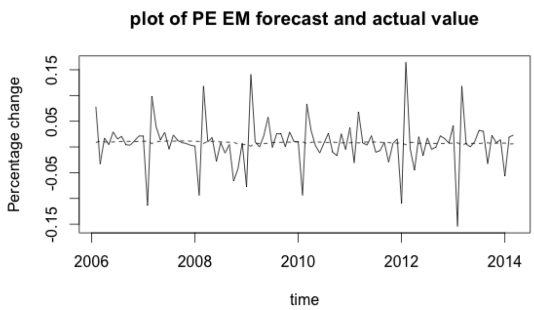
(c) Production of electricity



(d) CPI



(e) Industrial production



(f) Production of electricity

As exponential smoothing and FAVAR forecasts of CPI,industrial production and production of electricity significantly outperform the benchmark AR forecasts, this notable result suggest that the best approach to reduce forecasting error for six-month-ahead forecast is naive forecasting methodology.

5.4 Overall discussion and conclusion

The summary of performances of DI model and FARIMA model for the CPI, the industry production and the production of electricity can be seen in Table 5.4. Measured by relative MSFE, Table 5.4 presents number of DI and FARIMA forecasts that outperform, end with draw and underperform the benchmark AR forecasts[6]

Table 5.4: The number of diffusion index and factor-augmented autoregressive integrated moving average win, make a draw and lose to AR model for the CPI, the industry production and the production of electricity series

	Win	Draw	Loss
one-month-ahead	2	0	4
three-month-ahead	1	1	4
six-month-ahead	1	0	5
total	4	1	13

A first glance to table 5.4 reveals that total number of DI and FARIM forecasts that outperform AR forecasts is less than those of underperformed. The results of Table 5.4 that the static factor models that are estimated by PCA and have shown to have intrinsic value to forecast eight leading US macroeconomic in Stock and Watson (2002b) generally do not contribute forecasting improvements over benchmark AR forecasts for China's CPI, industrial production and production of electricity series. In most case, augmenting contemporaneous factors to AR and ARIMA model even produces additional forecasting errors. As for the PPI:industrial goods series, the DI and FARIMA model are useful to significatly improve the forecasting performances over benchmark AR forecasts.

The summary of performances of FAVAR model for four series is presented in table 5.5[7].

[6]There are 3 DI and 3 FARIMA models in each forecasting horizon so that total DI and FARIMA forecast are 24.

[7]There are 4 FAVAR forecasts in each forecasting horizon so that total forecasts are 12. The only 1 underperformed is FAVAR forecast of CPI at one-month-ahead

Table 5.5: The number of FAVAR wins, makes a draw and loses to AR by MSFE (four series)

	Win	Draw	Loss
MSFE	11	0	1
Total	11	0	1

As can be seen in Table 5.5, the number of FAVAR that outperform AR forecasts significantly outweigh to those of underperformed at 11-1. The inspection of Table 5.5 reveals a striking finding: simply using vector structure of lag factors and lag values captures vast majority of the forecasting improvements.

The performances of DI and FAVAR reverse the finding in Stock and Watson (2002b) but confirm Bernanke et al. (2004) and Fernald et al. (2014). These results might be explained by the concern about weak quality of China's data and dramatic change economic structure. As suggested by Bernanke et al. (2004), the FAVAR model is very appropriate when the economy is transforming its economic structure or the quality some of series in dataset might be unreliable (treated as latent variables), which is the case in this study. However, there is no evidence to show DI model is appropriate when some of series in dataset might be unreliable. Most of literature on applications of DI are Western countries where official data is generally reliable.

Nevertheless, it is important to note that even through the results of this study do not support the superiority of DI, it does not mean the DI is completely useless to forecasting China's leading variables. As Boivin and Ng (2005) claimed that "the composition of the data set and the dimensions of the cross-section are important in producing better forecasts from factor models." This arises the need of further research.

Another notable result is that the two winning forecasts in three-month-ahead and six-month-ahead: exponential smoothing forecast and FAVAR forecasts essentially generate naive forecasts for CPI, industrial production and production of electricity across out-of-sample forecasting periods (see Figure 5.2, 5.3 and 5.4) This is rather surprising result because it says the best model to forecast China's CPI, industrial production and production of electricity in next three and six months is just look at today's value. This raises a very interesting question: why the more complicated models do not improve forecasts over simpler models as the forecast horizon increases? It might be because Chinese statistical authorities have motivations to falsify the some of important macroeconomic data such as CPI and GDP in order to meet certain annual target and general public expectations. Also the margin of error in much of the published data is likely to be sufficiently large to allow the statistical authorities having a choice of final value from a relatively wide range of equally correct value.

Chapter 6

Conclusion

The focus of this study is to assess the forecasting performances of large dimensional approximate factor models for four China's leading macroeconomic variables - CPI, industrial production, production of electricity and PPI:industrial goods. The three factor models being used in this study are (1) DI that of Stock and Watson (2002b) which is augmenting static factor through PCA to standard AR model, (2) FARIMA which is augmenting static factor to a standard ARIMA model and (3) FAVAR which is dynamic structure of lag factors and lag variables. This study also assess the forecasting performances of ARIMA model and simple exponential smoothing model.

The factor forecasting is conducted into two steps. In step one, factors are estimated through PCA as suggested by Stock and Watson (2002b). This is because factors estimated by PCA is proxy of data in subsequent forecasting regression when N is sufficiently large, regardless ratio of N and T. The number of factor is selected by Bai and Ng (2002)'s information criteria to ensure the consistency between true number of factors and selected number of factors. Once factors are estimated and selected, they are used to forecast four leading variables in one-month, three-month and six-month-ahead. The lag length of AR, ARIMA, DI, FARIMA and FAVAR models are all selected recursively by AIC with maximum lag order of 24.

The full dataset contains 36 predictors representing an exhaustive description of China's economy including measure of government activity such as government revenue and government expenditure, real economic indicators such as industrial sales and production of electricity, financial indicators such as money supply and interest rate, and trade activity such as import and export. As suggested by Stock and Watson (2002b), the data are pre-processed in three stages before being modelled with a factor representation. In first stage, data are transformed into either difference or difference logarithms [1]. In second stage, transformed data are further standardised to have mean zero and unit variance. In final stage, the standardised data were screened for outliers.

Overall, the analysis of results suggests two findings. Firstly, the DI and FARIMA do not generaly improve forecasting accuracy except for PPI:industrial goods series in one-month, three-month and six-month ahead forecasts. Rather, the FAVAR model is a superior model to forecast four Chinese leading variables in one-,three-, and six-month-ahead. The favourable results of FAVAR model might be because of concerns about weak quality of China's official data and rapid change of economic structure.

Secondly, the naive forecast is a sensible approach to forecast the CPI, the industrial production and the production of electricity in the three-month and the six-month ahead forecasting horizons. This is a rather surprising result. It essentially implies that using complicated forecasting models such as DI and FARIMA forecasting methodology is unnecessary and might even generate extra forecasting errors to that of naive forecast.

[1] The decision to take difference or difference of logarithms was made judgmentally after preliminary data analysis, including inspection of time series plot of data and unit root test. In general, difference is taken for those already in index and percentage and difference of logarithm is taken for those are not in percentage change.

In terms of further research, it appears that the performances of forecasting models are sensitive to the model selection and the forecasting evaluation method. The relative MSFE is conventional forecasting evaluation method in many forecasting literature, but one question arising from using MSFE to measure forecasting performances is that even though DI forecasts generate the higher MSFE than that of AR forecasts (for instance DI forecast of industrial production generate 15% more forecasting error in one-month-ahead forecast), does it mean DI forecasting model have no intrinsic value to improve forecasting accuracy at all? Based on inspection of MSFE results in Table 5.2, a short answer to that question is "Yes", but the reason we have "Yes" is because DI performed extremely bad forecasts during global financial crisis. Indeed, it generates forecast that well capture the predictable dynamic during non-crisis period (the plot of it can be seen in appendix A). As a results, the under-performances in crisis period contribute considerable amount of forecasting error which out-performances in non-crisis period are insufficient to compensate. This particular example rises the need of more appropriate measurement of forecasting evaluation.

Furthermore, in what follows Schumacher and Dreger (2002) and Stock and Watson (2002b) they conducted a statistical test on whether the improvements from factor models are significant. Stock and Watson (2002b) found vast majority of improvements were statistically significant whereas Schumacher and Dreger (2002) found majority of improvements were insignificant. These conflicted results arise the question as to whether forecasting gains from FAVAR model are statistical significant for four China's leading macroeconomic variables.

There are five more possible ways to improve this study further. Firstly, this study only uses the balanced pool of data with same frequency. Stock and Watson (2002b) found that the performances of factor forecasts are generally better when unbalanced pool of data with mixed frequency are used. This findings rises possibility that further forecasting gains can be realized using unbalanced pool of data with mixed frequency. Secondly, some theoretical and applied literature (Lin and Wang, 2013) (Li, 1991) (Wold, 1985) developed and used different methodology to estimate factors such as sliced inverse regression and partial least square. These estimations of factors might potentially improve forecasting accuracy. Thirdly, the results of this study are based on 36 predictors chosen judgementally from large number of available macroeconomic time series. Would there be additional improvements if this study was to use 100 series or even more series? Fourthly, Marcellino et al. (2003) used a different data transformation technique to this study. They took difference of logarithms to all series whereas this study takes either difference or difference of logarithms. Also they seasonally adjusted all series whereas this study only seasonally CPI and production electricity. The different data transformation technique might produces different results. Finally, as the global financial crisis affected China greatly, mean GRAH model might be appropriate to model and forecasting China's leading variables in post crisis period.

Nevertheless, this study sheds light on China's macroeconomic forecasting in a data-rich environment which is essential for policy makers in China, exporters in Australia, investors who are extremely concern about future of China's economic condition and many more. Of particular importance is that this study contributes significance to existing literature in the sense that this is first study that uses large number of predictors to predict industrial production, electricity production and PPI:industrial goods which jointly reflect China's economic activity. This study also reaffirms the findings in (Fernald et al., 2014) and Bernanke et al. (2004) that FAVAR model is very appropriate when some of series in dataset might be unreliable and economic structure is rapidly changing, which is the case of China.

Appendices

Appendix A

Plot of DI and exponential smoothing forecasts

Figure A.1: DI forecasts of IP versus actual value in 3-month-ahead

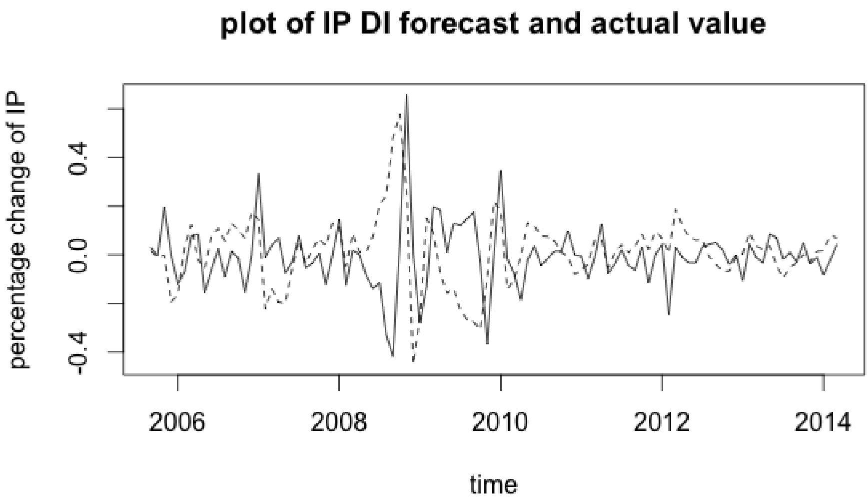
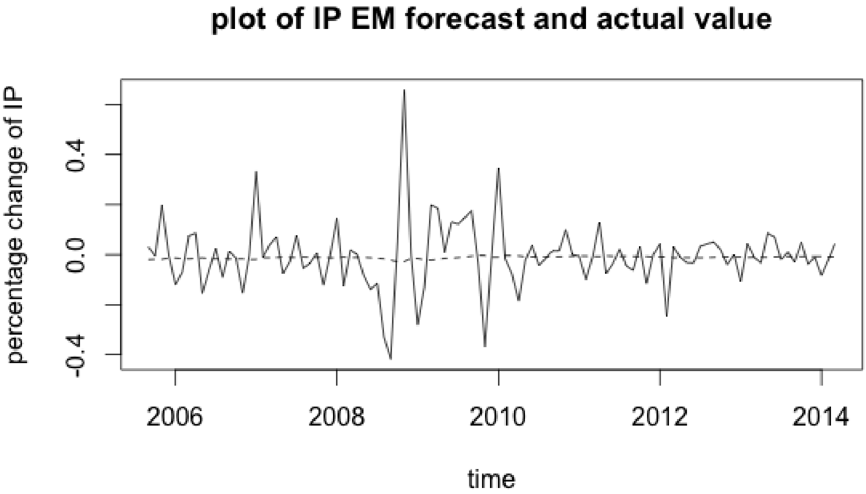


Figure A.2: Plot of exponential smoothing forecast versus actual for IP



By relative MSFE, the exponential smoothing forecast of the industrial production far more outweigh DI forecast; however, according to Figures above, the DI model produces relatively good forecasts during non-financial crisis period while the exponential smoothing generates a naive forecasts. This arises the need of more appropriate forecasting evaluation methods.

Appendix B

Relativel mean absoulte forecasting error

The relativel mean absolute forecasting error is another popular forecasting performance evaluation used in forecasting literature. Similar to relative MSFE, relative MAFE is computed by setting up the MAFE of AR to be 1. Essentially, relative MAFE produced very similar results those of relative MSFE. Tables below reveal the relative MAFE for 1-month, 3-month and 6-month ahead forecasting.

Table B.1: Relative mean absoulte forecasting errors for one-month-ahead forecast

	CPI	IP	PE	PPI
AR	1.00	1.00	1.00	1.00
ARIMA	0.93	1.14	1.01	0.85
ES	1.26	0.67	0.96	1.11
DI	1.09	1.07	1.02	0.91
FARIMA	0.94	1.11	1.14	0.90
FAVAR	1.12	0.86	0.96	0.91

As can be seen in the table, DI and FARIMA models do not general improve the forecasting performances; however the FAVAR model yield moderate improvement over benchmark AR model for the IP, PE and PPI series.

Table B.2: Relative mean absoulte forecasting errors for three-month-ahead forecast

	CPI	IP	PE	PPI
AR	1.00	1.00	1.00	1.00
ARIMA	0.82	1.28	1.10	0.95
ES	0.81	0.80	0.93	0.86
DI	1.00	1.06	1.02	0.79
FARIMA	0.83	1.29	1.08	0.75
FAVAR	0.78	0.85	0.91	0.85

According to the table, ES and FAVAR forecasts yield some huge improvements over the benchmark AR forecasts; however the DI and FARIMA generally perform worse than benchmark AR forecasts.

Table B.3: Relative mean absoulte forecasting errors for six-month-ahead forecast

	CPI	IP	PE	PPI
AR	1.00	1.00	1.00	1.00
ARIMA	0.89	1.17	1.15	0.96
ES	0.79	0.85	0.98	0.71
DI	0.99	1.00	1.02	0.65
FARIMA	0.90	1.16	1.12	0.65
FAVAR	0.79	0.85	0.98	0.84

The results of relative MAFE for six-month-ahead forecast reveal a similar pattern to the one for three-month-ahead. More specifically, the ES and FAVAR still to produce superior forecasts than benchmark AR forecasts while the DI and FARIMA generally produced inferior forecasts.

Appendix C

R code

This is R script I have done to generate results. The R packages required are seasonal (X-13), var, tseries,forecast and ggplot2. Note that seasonal (X-13) package is unavailable in Mac so I use my friend's PC to generate seasonally adjusted CPI and PE and then copy to my mac. All codes and data are upon requested.

```
ADcpi<- ts(adcpi,start=1997,frequency=12)
dim(ADcpi)
plot(ADcpi, main="seasonally adjusted month-on-month CPI",
xlab="time", ylab="%")
adf.test(ADcpi)
China'scpi <- ts(China's.CPI.MoM, start=1997,frequency=12)
plot(China'scpi)
acf(China'scpi,lag.max=24)
acf(ADcpi, lag.max=24)
dADcpi <- diff(ADcpi) # take first difference
plot(dADcpi, main="seasonally adjusted China's monthly inflation rate"
, xlab="time", ylab="percentage", col="blue")
# now fit to ar model first
```

```
dADcpi1 <- ts(dADcpi[1:103],start=c(1997,02),frequency=12)
model1 <- ar(dADcpi1,method="ols",aic=TRUE)
model1
predict(model1,n.ahead=1)
# now doing a loop
T <- 205
start <- 103
forecasts <- NA
orders <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  model <- ar(CPIs, method="ols", aic=TRUE)
  orders[i] <- model$order
  forecastm <- predict(model, n.ahead=1)
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
orderar <- ts(orders, start=c(1997,02), frequency=12)
cpiar <- ts(forecasts, start=c(1997,02), frequency=12)
cpiar #this is AR forecast for CPI
orderar # this is AR order that we need to use in later factor model
# now doing arima for AR
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  model <- auto.arima(CPIs,ic=c("aic"))
  forecastm <- predict(model, n.ahead=1)
```

```
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
cpiarima <- ts(forecasts, start=c(1997,02), frequency=12)
cpiarima
# now doing a exponential smoothing model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  model <- HoltWinters(CPIs, beta=FALSE,gamma=FALSE)
  forecastm <- forecast(model, h=1)
  forecastmt <- matrix(forecastm$mean)
  forecasts[i+1] <- forecastmt[1,1]
}
cpies <- ts(forecasts, start=c(1997,02), frequency=12)
cpies

# now doing a factor model
totaldatacpi<- ts(paper.data.editedCPI,start=1997,frequency=12)
class(totaldatacpi)
dim(totaldatacpi)
a <- diff(log(totaldatacpi[,1:22]))
dim(a)
head(a)
b<- diff(totaldatacpi[,23:35])
head(b)
dtotaldata <- cbind(a,b)
```

```
dim(dtotaldata)
head(dtotaldata)
dtotaledited <- scale(dtotaldata, center=TRUE, scale=TRUE)
class(dtotaldata)
dtotaledited
class(dtotaledited)
dim(dtotaledited)
dtotaledited1 <- ts(dtotaledited[1:103,],
start=c(1997,02),frequency=12)
dim(dtotaledited1)
pca1 <- princomp(dtotaledited1, cor=TRUE)
summary(pca1)
class(pca1$scores)
dim(pca1$scores)
POETKhat(dtotaledited1)
factor1 <- ts(pca1$scores[1:103,1:1], start=c(1997,02),frequency=12)
factor1
class(factor1) # it is time series and ready to fit DI model
modelDI1 <- arima(dADcpi1,order=c(orderar[103],0,0),xreg=factor1)
modelDI1
forecastDI1 <- predict(modelDI1,n.ahead=1,newxreg=factor1[103])
forecastDI1
modelDI2 <-auto.arima(dADcpi1,d=NA,D=NA,
max.p=24,max.q=0,max.P=0,max.Q=0,
max.d=0,max.D=0,ic=c("aic"), xreg=factor1)
modelDI2
forecastDI2 <- predict(modelDI2,n.ahead=1,newxreg=factor1[103])
forecastDI2
# now rolling it up
```

```
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  dtotalediteds <- ts(dtotaledited[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedts, cor=TRUE)
  m<- POETKhat(dtotaleditedts)
  factors <- ts(pca1$scores[1:i,1:m$K1BN], start=c(1997,02),frequency=12)
  model <- auto.arima(CPIs,d=NA,D=NA,
  max.p=24,max.q=0,max.P=0,max.Q=0,max.
  d=0,max.D=0,ic=c("aic"), xreg=factors)
  forecastm <- predict(model, n.ahead=1,newxreg=factors[i])
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
dicpi <- ts(forecasts, start=c(1997,02), frequency=12)
dicpi
# now adding factors to arima

T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedts <- ts(dtotaledited[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedts, cor=TRUE)
```

```
  m<- POETKhat(dtotaleditedS)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(CPIs,xreg=factors)
  forecastm <- predict(model, n.ahead=1,newxreg=factors[i])
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
FARIMAcpi <- ts(forecasts, start=c(1997,02), frequency=12)
FARIMAcpi
# trying to do a FAVAR approach
pca1 <- princomp(dtotaledited1, cor=TRUE)
summary(pca1)
class(pca1$scores)
dim(pca1$scores)
POETKhat(dtotaledited1)
factor1 <- ts(pca1$scores[1:103,1], start=c(1997,02),frequency=12)  #
factor1
varobject <-cbind(dADcpi1,factor1)
class(varobject)
head(varobject)
dim(varobject)
plot(varobject)
var1 <- VAR(varobject,p=1,type=c("const"),ic=c("AIC"))
forecast1=forecast(var1,h=1)
forecast1$mean$dADcpi1
# now doing var loop
T <- 205
start <- 103
```

```
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedS <- ts(dtotaledited[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedS, cor=TRUE)
  m<- POETKhat(dtotaleditedS)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  varcpidata <- cbind(CPIs,factors)
  model <- VAR(varcpidata,p=1,type=c("const"),ic=c("AIC"))
  forecastm <- forecast(model,h=1)
  forecastmt <- forecastm$mean$CPIs
  forecasts[i+1] <- forecastmt
}
FAVARcpi <- ts(forecasts, start=c(1997,02), frequency=12)
FAVARcpi

#now doing IP
tIP <- ts(adip,start=1997,frequency=12)
class(tIP)
dim(tIP)
plot(tIP, main="China's monthly seasonal adjusted industrial
production", xlab="time", ylab="billion Yuan")
adf.test(tIP)
lIP=log(tIP)
lIP
ldIP=diff(lIP)
class(ldIP[1:103])
```

```
IP1 <- ts(ldIP[1:103,],start=c(1997,02),frequency=12)
class(IP1)
# now doing a loop for AR
T <- 205
start <- 103
forecasts2 <- NA
orders2 <- NA
for (i in start:T){
  IPc<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  model <- ar(IPc, method="ols", aic=TRUE)
  orders2[i] <- model$order
  forecastm <- forecast(model, h=1)
  forecastmt <- matrix(forecastm$mean)
  forecasts2[i+1] <- forecastmt[1,1]
}
orderip <- ts(orders2, start=c(1997,02), frequency=12)
arip <- ts(forecasts2, start=c(1997,02), frequency=12)
orderip
arip

# doing arima loop
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  IPc<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  model <- auto.arima(IPc,ic=c("aic"))
  forecastm <- forecast(model, h=1)
  forecastmt <- matrix(forecastm$mean)
```

```
  forecasts[i+1] <- forecastmt[1,1]
}
iparima <- ts(forecasts, start=c(1997,02), frequency=12)
iparima

#finally doing exponential smoothing for IP
T <- 205
start <- 103
forecasts <- NA
orders <- NA
for (i in start:T){
  dtc<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  model <- HoltWinters(dtc, beta=FALSE,gamma=FALSE)
  forecastm <- forecast(model, h=1)
  forecastmt <- matrix(forecastm$mean)
  forecasts[i+1] <- forecastmt[1,1]
}
ipes <- ts(forecasts, start=c(1997,02), frequency=12)
ipes

# now doing diffusion index
totaldataIP<- ts(paper.data.editedIP,start=1997,frequency=12)
class(totaldataIP)
c <- diff(log(totaldataIP[,1:21]))
dim(c)
d <- diff(totaldataIP[,22:35])
head(d)
dtotaldataIP <- cbind(c,d)
dtotaleditedIP <- scale(dtotaldataIP, center=TRUE, scale=TRUE)
```

```
dtotaleditedIP
class(dtotaleditedIP)
dim(dtotaleditedIP)
dtotaledited1IP <- ts(dtotaleditedIP[1:103,],
start=c(1997,02),frequency=12)
dim(dtotaledited1IP)
pca1IP <- princomp(dtotaledited1IP, cor=TRUE)
summary(pca1IP)
class(pca1IP$scores)
dim(pca1IP$scores)
POETKhat(dtotaledited1IP)
factor1IP <- ts(pca1IP$scores[1:103,1],
start=c(1997,02),frequency=12)
class(factor1IP)
modelDI1IP <- arima(IP1,order=c(12,0,0),xreg=factor1IP)
forecastDI1IP <- predict(modelDI1IP,n.ahead=1,newxreg=factor1IP[103])
forecastDI1IP
# now rolling it up
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  IPs<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedS <-
  ts(dtotaleditedIP[1:i,],start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedS, cor=TRUE)
  m <- POETKhat(dtotaleditedS)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
```

```
  model <- auto.arima(IPs,d=NA,D=NA,max.p=24,
  max.q=0,max.P=0,max.Q=0,max.d=0,
  max.D=0,ic=c("aic"), xreg=factors)
  forecastm <- predict(model, n.ahead=1,newxreg=factors[i])
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
diip <- ts(forecasts, start=c(1997,02), frequency=12)
diip
# now adding factor model to arima
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  IPs<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  dtotaledited <-ts(dtotaleditedIP[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaledited, cor=TRUE)
  m <- POETKhat(dtotaledited)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(IPs,xreg=factors)
  forecastm <- predict(model, n.ahead=1,newxreg=factors[i])
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
FARIMAip <- ts(forecasts, start=c(1997,02), frequency=12)
FARIMAip
# now doing a FAVAR for ip
```

```
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  IPs<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedS <- ts(dtotaleditedIP[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedS, cor=TRUE)
  m <- POETKhat(dtotaleditedS)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  varipdata <- cbind(IPs,factors)
  model <- VAR(varipdata,p=1,type=c("const"),ic=c("AIC"))
  forecastm <- forecast(model,h=1)
  forecastmt <- forecastm$mean$IPs
  forecasts[i+1] <- forecastmt
}
FAVARIP <- ts(forecasts, start=c(1997,02), frequency=12)
FAVARIP
# now doing a electricity production
adPE <- ts(adpe,start=1997,frequency=12)
dim(adPE)
plot(adPE)
adf.test(adPE)
dladPE <- diff(log(adPE)) # take first difference
plot(dladPE, main="percentage change of prodcution of electricity",
xlab="time", ylab="percentage", col="blue")
plot(adPE, main="volume of production of electricity",
xlab="time", ylab="volume")
```

```
# now fit to ar model first
dladPE1 <- ts(dladPE[1:103],start=c(1997,02),frequency=12)
model1 <- ar(dladPE,method="ols",aic=TRUE)
model1
predict(model1,n.ahead=1)
# now doing a loop
T <- 205
start <- 103
forecasts <- NA
orders <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  model <- ar(PEs, method="ols", aic=TRUE)
  orders[i] <- model$order
  forecastm <- predict(model, n.ahead=1)
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
orderpe <- ts(orders, start=c(1997,02), frequency=12)
pear <- ts(forecasts, start=c(1997,02), frequency=12)
pear
orderpe
# now doing arima for AR
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  model <- auto.arima(PEs,ic=c("aic"))
```

```
  forecastm <- predict(model, n.ahead=1)
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
pearima <- ts(forecasts, start=c(1997,02), frequency=12)
pearima
# now doing a exponential smoothing model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  model <- HoltWinters(PEs, beta=FALSE,gamma=FALSE)
  forecastm <- forecast(model, h=1)
  forecastmt <- matrix(forecastm$mean)
  forecasts[i+1] <- forecastmt[1,1]
}
pees <- ts(forecasts, start=c(1997,02), frequency=12)
pees

# now doing a factor model
totaldatape<- ts(paper.data.editedPE,start=1997,frequency=12)
class(totaldatape)
e <- diff(log(totaldatacpi[,1:21]))
dim(e)
head(e)
f<- diff(totaldatacpi[,22:35])
head(f)
dtotaldatape <- cbind(e,f)
```

```
dtotaleditedpe <- scale(dtotaldatape, center=TRUE, scale=TRUE)
dtotaleditedpe
class(dtotaleditedpe)
dim(dtotaleditedpe)
dtotaleditedpe1 <- ts(dtotaleditedpe[1:103,],
start=c(1997,02),frequency=12)
dim(dtotaleditedpe1)
pca1 <- princomp(dtotaleditedpe1, cor=TRUE)
summary(pca1)
class(pca1$scores)
dim(pca1$scores)
POETKhat(dtotaleditedpe1)
factor1 <- ts(pca1$scores[1:103,1], start=c(1997,02),frequency=12)  #
factor1
class(factor1) # it is time series and ready to fit DI model
modelDI1 <- arima(dladPE1,order=c(orderpe[103],0,0),xreg=factor1)
forecastDI1 <- predict(modelDI1,n.ahead=1,newxreg=factor1[103])
forecastDI1
# now rolling it up
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedpes <-
  ts(dtotaleditedpe[1:i,],start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedpes, cor=TRUE)
  m<- POETKhat(dtotaleditedpes)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
```

```
  start=c(1997,02),frequency=12)
  model <- auto.arima(PEs,d=NA,D=NA,max.p=24,max.q=0,max.P=0,
  max.Q=0,max.d=0,max.D=0,ic=c("aic"), xreg=factors)
  forecastm <- predict(model, n.ahead=1,newxreg=factors[i])
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
dipe <- ts(forecasts, start=c(1997,02), frequency=12)
dipe
# now adding the factor to arima model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedpes <- ts(dtotaleditedpe[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedpes, cor=TRUE)
  m<- POETKhat(dtotaleditedpes)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(PEs,xreg=factors)
  forecastm <- predict(model, n.ahead=1,newxreg=factors[i])
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
FARIMApe <- ts(forecasts, start=c(1997,02), frequency=12)
FARIMApe
# now doing FAVAR for PE
```

```
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedpes <- ts(dtotaleditedpe[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedpes, cor=TRUE)
  m<- POETKhat(dtotaleditedpes)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  varipdata <- cbind(PEs,factors)
  model <- VAR(varipdata,p=1,type=c("const"),ic=c("AIC"))
  forecastm <- forecast(model,h=1)
  forecastmt <- forecastm$mean$PEs
  forecasts[i+1] <- forecastmt
}
FAVARpe <- ts(forecasts, start=c(1997,02), frequency=12)
FAVARpe

# now doing producer price index
adPPI <- ts(adppi,start=1997,frequency=12)
dim(adPPI )
plot(adPPI )
adf.test(adPPI )
dadPPI  <- diff(adPPI ) # take first difference
plot(dadPPI , main="first difference of producer price index",
xlab="time", ylab="%change", col="blue")
plot(adPPI, main="adjusted monthly producer price index",
```

```
xlab="time", ylab="%")
# now fit to ar model first
dadPPI1 <- ts(dadPPI[1:103],start=c(1997,02),frequency=12)
model1 <- ar(dadPPI1,method="mle",aic=TRUE)
model1
predict(model1,n.ahead=1)
# now doing a loop
T <- 205
start <- 103
forecasts <- NA
orders <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  model <- ar(PPIs, method="mle", aic=TRUE)
  orders[i] <- model$order
  forecastm <- predict(model, n.ahead=1)
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
orderppi <- ts(orders, start=c(1997,02), frequency=12)
ppiar <- ts(forecasts, start=c(1997,02), frequency=12)
ppiar #this is AR forecast for CPI
orderppi # this is AR order that we need to use in later factor model
# now doing arima for AR
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
```

```
  model <- auto.arima(PPIs,ic=c("aic"))
  forecastm <- predict(model, n.ahead=1)
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
ppiarima <- ts(forecasts, start=c(1997,02), frequency=12)
ppiarima
# now doing a exponential smoothing model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  model <- HoltWinters(PPIs, beta=FALSE,gamma=FALSE)
  forecastm <- forecast(model, h=1)
  forecastmt <- matrix(forecastm$mean)
  forecasts[i+1] <- forecastmt[1,1]
}
ppies <- ts(forecasts, start=c(1997,02), frequency=12)
ppies

# now doing a factor model
totaldatappi<- ts(paper.data.editedPPI,start=1997,frequency=12)
class(totaldatappi)
g <- diff(log(totaldatacpi[,1:22]))
dim(g)
head(g)
h<- diff(totaldatacpi[,23:35])
head(h)
```

```
dtotaldatappi <- cbind(g,h)
dim(totaldatappi)
dtotaldatappied <- scale(dtotaldatappi, center=TRUE, scale=TRUE)
class(dtotaldatappied)
dim(dtotaldatappied)
dtotaldatappied1 <- ts(dtotaldatappied[1:103,],
start=c(1997,02),frequency=12)
dim(dtotaldatappied1)
pca1 <- princomp(dtotaldatappied1, cor=TRUE)
summary(pca1)
class(pca1$scores)
dim(pca1$scores)
POETKhat(dtotaldatappied1)
factor1 <- ts(pca1$scores[1:103,1], start=c(1997,02),frequency=12)  #
factor1
class(factor1)
modelDI1 <- arima(dadPPI1,order=c(orderpe[103],0,0),xreg=factor1)
forecastDI1 <- predict(modelDI1,n.ahead=1,newxreg=factor1[103])
forecastDI1
# now rolling it up
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  dtotaldatappieds <- ts(dtotaldatappied[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaldatappieds, cor=TRUE)
  m<- POETKhat(dtotaldatappieds)
```

```
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(PPIs,d=NA,D=NA,max.p=24,max.q=0,
  max.P=0,max.Q=0,max.d=0,max.D=0,ic=c("aic"), xreg=factors)
  forecastm <- predict(model, n.ahead=1,newxreg=factors[i])
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
dippi <- ts(forecasts, start=c(1997,02), frequency=12)
dippi
# now adding the factors to arima model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  dtotaldatappieds <- ts(dtotaldatappied[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaldatappieds, cor=TRUE)
  m<- POETKhat(dtotaldatappieds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(PPIs,xreg=factors)
  forecastm <- predict(model, n.ahead=1,newxreg=factors[i])
  forecastmt <- forecastm$pred
  forecasts[i+1] <- forecastmt
}
FARIMAppi <- ts(forecasts, start=c(1997,02), frequency=12)
FARIMAppi
```

```
# now doing FAVAR for PPI
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  dtotaldatappieds <- ts(dtotaldatappied[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaldatappieds, cor=TRUE)
  m<- POETKhat(dtotaldatappieds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  varipdata <- cbind(PPIs,factors)
  model <- VAR(varipdata,p=1,type=c("const"),ic=c("AIC"))
  forecastm <- forecast(model,h=1)
  forecastmt <- forecastm$mean$PPIs
  forecasts[i+1] <- forecastmt
}
FAVARppi <- ts(forecasts, start=c(1997,02), frequency=12)
FAVARppi
# now writing user defined functions to compute MSE and AME
mse <- function(x,y,n){
  sum((x-y)^2)/n
}
n <- 103 # this is total number of out-of-samle period

cpiAR <- mse(dADcpi[104:206],cpiar[104:206],n)
cpiARIMA <- mse(dADcpi[104:206],cpiarima[104:206],n)
cpiES <- mse(dADcpi[104:206],cpies[104:206],n)
```

```
cpiDI <- mse(dADcpi[104:206],dicpi[104:206],n)
cpiFARIMA <- mse(dADcpi[104:206],FARIMAcpi[104:206],n)
cpiFAVAR <- mse(dADcpi[104:206],FAVARcpi[104:206],n)
RcpiAR <-cpiAR/cpiAR
RcpiARIMA <-cpiARIMA/cpiAR
RcpiES=cpiES/cpiAR
RcpiDI=cpiDI/cpiAR
RcpiFARIMA=cpiFARIMA/cpiAR
RcpiFAVAR=cpiFAVAR/cpiAR
CPIforecastMSE <-cbind(RcpiAR,RcpiARIMA,RcpiES,
RcpiDI,RcpiFARIMA,RcpiFAVAR)
CPIforecastMSE
tsdADcpi <- ts(dADcpi[104:206],start=c(2005,09),frequency=12)
tscpiar <- ts(cpiar[104:206],start=c(2005,09),frequency=12)
ts.plot(tsdADcpi,tscpiar,
        gpars=list(main="plot of CPI AR forecast and actual
        value",xlab="time",ylab="differecen of CPI",lty=c(1:2)))
tscpiarima <-ts(cpiarima[104:206],start=c(2005,09),frequency=12)
ts.plot(tsdADcpi,tscpiarima,
        gpars=list(main="plot of CPI ARIMA forecast and actual
        value",xlab="time",ylab="differecen of CPI",lty=c(1:2)))
tscpies <-ts(cpies[104:206],start=c(2005,09),frequency=12)
ts.plot(tsdADcpi,tscpies,
        gpars=list(main="plot of CPI EM forecast and actual
        value",xlab="time",ylab="differecen of CPI",lty=c(1:2)))
tsdicpi <-ts(dicpi[104:206],start=c(2005,09),frequency=12)
ts.plot(tsdADcpi,tsdicpi,
       gpars=list(main="plot of CPI DI forecast and actual
       value",xlab="time",ylab="differecen of CPI",lty=c(1:2)))
```

```
tsFARIMAcpi <-ts(FARIMAcpi[104:206],start=c(2005,09),frequency=12)
ts.plot(tsdADcpi,tsFARIMAcpi,
        gpars=list(main="plot of CPI FARIMA forecast and actual
        value",xlab="time",ylab="differecen of CPI",lty=c(1:2)))
tsFAVARcpi <-ts(FAVARcpi[104:206],start=c(2005,09),frequency=12)
ts.plot(tsdADcpi,tsFAVARcpi,
        gpars=list(main="plot of CPI FAVAR forecast and actual
        value",xlab="time",ylab="differecen of CPI",lty=c(1:2)))
# now for IP
ipAR <- mse(ldIP[104:206],arip[104:206],n)
ipARIMA <- mse(ldIP[104:206],iparima[104:206],n)
ipES <- mse(ldIP[104:206],ipes[104:206],n)
ipDI <- mse(ldIP[104:206],diip[104:206],n)
ipFARIMA <- mse(ldIP[104:206],FARIMAip[104:206],n)
ipFAVAR <-mse(ldIP[104:206],FAVARIP[104:206],n)
IPforecastMSE <-cbind((ipAR\ipAR),(ipARIMA\ipAR),
(ipES\\ipAR),(ipDI\ipAR),(ipFARIMA\ipAR),(ipFAVAR\ipAR)
IPforecastMSE
tsIP <- ts(ldIP[104:206],start=c(2005,09),frequency=12)
tsarip <- ts(arip[104:206],start=c(2005,09),frequency=12)
ts.plot(tsIP,tsarip,
        gpars=list(main="plot of IP AR forecast and actual
        value",xlab="time",ylab="percentage change of
        IP",lty=c(1:2)))
tsiparima <-ts(iparima[104:206],start=c(2005,09),frequency=12)
ts.plot(tsIP,tsiparima,
        gpars=list(main="plot of IP ARIMA forecast and actual
        value",xlab="time",ylab="percentage change of
        IP",lty=c(1:2)))
```

```
tsipes <-ts(ipes[104:206],start=c(2005,09),frequency=12)
ts.plot(tsIP,tsipes,
        gpars=list(main="plot of IP EM forecast and actual
        value",xlab="time",ylab="percentage change of
        IP",lty=c(1:2)))
tsdiip <-ts(diip[104:206],start=c(2005,09),frequency=12)
ts.plot(tsIP,tsdiip,
        gpars=list(main="plot of IP DI forecast and actual
        value",xlab="time",ylab="percentage change of
        IP",lty=c(1:2)))
tsFARIMAip <-ts(FARIMAip[104:206],start=c(2005,09),frequency=12)
ts.plot(tsIP,tsFARIMAip,
        gpars=list(main="plot of IP FARIMA forecast and actual
        value",xlab="time",ylab="percentage change of
        IP",lty=c(1:2)))
tsFAVARip <-ts(FAVARIP[104:206],start=c(2005,09),frequency=12)
ts.plot(tsIP,tsFAVARip,
        gpars=list(main="plot of PE FAVAR forecast and actual
        value",xlab="time",ylab="percentage change of
        IP",lty=c(1:2)))
# now for Production of electricity
peAR <- mse(dladPE[104:206],pear[104:206],n)
peARIMA <- mse(dladPE[104:206],pearima[104:206],n)
peES <- mse(dladPE[104:206],pees[104:206],n)
peDI <- mse(dladPE[104:206],dipe[104:206],n)
peFARIMA <- mse(dladPE[104:206],FARIMApe[104:206],n)
peFAVAR <- mse(dladPE[104:206],FAVARpe[104:206],n)
PEforecastMSE<-cbind(peAR,peARIMA,peES,peDI,peFARIMA,peFAVAR)
tsPE <- ts(dladPE[104:206],start=c(2005,09),frequency=12)
```

```
tspear <- ts(pear[104:206],start=c(2005,09),frequency=12)
ts.plot(tsPE,tspear,
        gpars=list(main="plot of PE AR forecast and actual
        value",xlab="time",ylab="percentage change of
        PE",lty=c(1:2)))
tspearima <-ts(pearima[104:206],start=c(2005,09),frequency=12)
ts.plot(tsPE,tspearima,
        gpars=list(main="plot of PE ARIMA forecast and actual
        value",xlab="time",ylab="percentage change of
        PE",lty=c(1:2)))
tspees <-ts(pees[104:206],start=c(2005,09),frequency=12)
ts.plot(tsPE,tspees,
        gpars=list(main="plot of PE EM forecast and actual
        value",xlab="time",ylab="percentage change of
        PE",lty=c(1:2)))
tsdipe <-ts(dipe[104:206],start=c(2005,09),frequency=12)
ts.plot(tsPE,tsdipe,
        gpars=list(main="plot of PE DI forecast and actual
        value",xlab="time",ylab="percentage change of
        PE",lty=c(1:2)))
tsFARIMApe <-ts(FARIMApe[104:206],start=c(2005,09),frequency=12)
ts.plot(tsPE,tsFARIMApe,
        gpars=list(main="plot of PE FARIMA forecast and actual
        value",xlab="time",ylab="percentage change of
        PE",lty=c(1:2)))
tsFAVARpe <-ts(FAVARpe[104:206],start=c(2005,09),frequency=12)
ts.plot(tsPE,tsFAVARpe,
        gpars=list(main="plot of PE FAVAR forecast and actual
        value",xlab="time",ylab="percentage change of
```

```
        PE",lty=c(1:2)))

#now for PPI
ppiAR <- mse(dadPPI[104:206],ppiar[104:206],n)
ppiARIMA <- mse(dadPPI[104:206],ppiarima[104:206],n)
ppiES <- mse(dadPPI[104:206],ppies[104:206],n)
ppiDI <- mse(dadPPI[104:206],dippi[104:206],n)
ppiFARIMA <- mse(dadPPI[104:206],FARIMAppi[104:206],n)
ppiFAVAR <- mse(dadPPI[104:206],FAVARppi[104:206],n)
PPIforecastMSE <- cbind(ppiAR,ppiARIMA,ppiES,
ppiDI,ppiFARIMA,ppiFAVAR)
CPIforecastMSE
IPforecastMSE
PEforecastMSE
PPIforecastMSE
totalforecast <-rbind(CPIforecastMSE,IPforecastMSE,PEforecastMSE,PPIforecastMSE)
totalforecast
colnames(totalforecast) <- c("AR","ARIMA","ES","DI","FARIMA","FAVAR")
rownames(totalforecast) <- c("CPI","IP","PE","PPI")
totalMSE <-t(totalforecast)
totalMSE
tsPPI <- ts(dadPPI[104:206],start=c(2005,09),frequency=12)
tsppiar <- ts(ppiar[104:206],start=c(2005,09),frequency=12)
ts.plot(tsPPI,tsppiar,
        gpars=list(main="plot of PPI AR forecast and actual
        value",xlab="time",ylab="difference of PPI",lty=c(1:2)))
tsppiarima <-ts(ppiarima[104:206],start=c(2005,09),frequency=12)
ts.plot(tsPPI,tsppiarima,
        gpars=list(main="plot of PPI ARIMA forecast and actual
```

```
        value",xlab="time",ylab="difference of PPI",lty=c(1:2)))
tsppies <-ts(ppies[104:206],start=c(2005,09),frequency=12)
ts.plot(tsPPI,tsppies,
        gpars=list(main="plot of PPI EM forecast and actual
        value",xlab="time",ylab="difference of PPI",lty=c(1:2)))
tsdippi <-ts(dippi[104:206],start=c(2005,09),frequency=12)
ts.plot(tsPPI,tsdippi,
        gpars=list(main="plot of PPI DI forecast and actual
        value",xlab="time",ylab="difference of PPI",lty=c(1:2)))
tsFARIMAppi <-ts(FARIMAppi[104:206],start=c(2005,09),frequency=12)
ts.plot(tsPPI,tsFARIMAppi,
        gpars=list(main="plot of PPI FARIMA forecast and actual
        value",xlab="time",ylab="difference of PPI",lty=c(1:2)))
tsFAVARppi <-ts(FAVARppi[104:206],start=c(2005,09),frequency=12)
ts.plot(tsPPI,tsFAVARppi,
        gpars=list(main="plot of PPI FAVAR forecast and actual
        value",xlab="time",ylab="difference of PPI",lty=c(1:2)))


totalMSE
totalMSE3
totalMSE6
# now doing relatively MSE and MAE
RCPImse <- totalMSE[,1]/totalMSE[1,1]
RIPmse <- totalMSE[,2]/totalMSE[1,2]
RPEmse <- totalMSE[,3]/totalMSE[1,3]
RPPImse <-totalMSE[,4]/totalMSE[1,4]
RCPImse
RIPmse
```

```
RPEmse
RPPImse
rbind(RCPImse,RIPmse,RPEmse,RPPImse)
RtotalMSE <- t(rbind(RCPImse,RIPmse,RPEmse,RPPImse))
colnames(RtotalMSE) <- c("CPI","IP","PE","PPI")
RtotalMSE

totalMAFE <- xtable(RtotalMAE)
totalMSFE <- xtable(RtotalMSE)
print.xtable(totalMAFE,type="latex", file="",floating=TRUE,table.placement="H")
print.xtable(totalMSFE,type="latex", file="",floating=TRUE,table.placement="H")

tsIP <- ts(IP,start=1997,frequency=12)
tsCPI <- ts(China's.CPI.MoM, start=1997,frequency=12)
tsPE <- ts(production.electrictity, start=1997,frequency=12)
tsPPI <- ts(Producer.Price.Index..Industrial.Products,
start=1997,frequency=12)
plot(tsIP, main="Monthly Volume of China's Industrial Production",
ylab="Billion Yuan")
plot(tsCPI, main="Month-on-Month China's CPI last month=100",
ylab="%", col="red")
plot(tsPE, main="Monthly Volume of China's Electricity
production",ylab="KWH Billion",col="blue")
plot(tsPPI, main="Monthly PPI:Industrial Products last year=100",
ylab="%")
plot(adPE, main="Seasonally Adjusted Production of
Electricity",ylab="KWH Billion")
rownames(resultsummary)<-
c("one-month-ahead","three-month-ahead","six-month-ahead","total")
```

```
MSFEresult <- resultsummary[1:4,1:3]
MAFEresult <- resultsummary[1:4,4:6]
colnames(MAFEresult) <-c("Win","Draw","Lose")
xMSFEresult <-xtable(MSFEresult)
xMAFEresult <-xtable(MAFEresult)
print.xtable(xMSFEresult,type="latex",
file="",floating=TRUE,table.placement="H")
print.xtable(xMAFEresult,type="latex",
file="",floating=TRUE,table.placement="H")

#Three month ahead forecasts
ADcpi<- ts(adcpi,start=1997,frequency=12)
dim(ADcpi)
plot(ADcpi, main="month-on-month CPI", xlab="time", ylab="%")
adf.test(ADcpi)
China'scpi <- ts(China's.CPI.MoM, start=1997,frequency=12)
plot(China'scpi)
acf(China'scpi,lag.max=24)
acf(ADcpi, lag.max=24)
dADcpi <- diff(ADcpi) # take first difference
plot(dADcpi, main="seasonally adjusted China's monthly inflation
rate", xlab="time", ylab="percentage", col="blue")
# now fit to ar model first
dADcpi1 <- ts(dADcpi[1:103],start=c(1997,02),frequency=12)
model1 <- ar(dADcpi1,method="ols",aic=TRUE)
model1
predict(model1,n.ahead=3)
forecast3month <- predict(model1,n.ahead=3)$pred[3]
forecast3month
```

```
dADcpi2 <- ts(dADcpi[1:106],start=c(1997,02),frequency=12)
model2<- ar(dADcpi2,method="ols",aic=TRUE)
predict(model2,n.ahead=3)


# now doing a loop
T <- 205
start <- 103
forecasts <- NA
orders <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  model <- ar(CPIs, method="ols", aic=TRUE)
  orders[i] <- model$order
  forecastm <- predict(model, n.ahead=3)
  forecastmt <- forecastm$pred[3]
  forecasts[i+1] <- forecastmt
}
orderar3 <- ts(orders, start=c(1997,02), frequency=12)
cpiar3 <- ts(forecasts, start=c(1997,04), frequency=12)
cpiar3 #this is AR forecast for CPI
orderar3 # this is AR order that we need to use in later factor model
# now doing arima for AR
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  model <- auto.arima(CPIs,ic=c("aic"))
  forecastm <- predict(model, n.ahead=3)
```

```
  forecastmt <- forecastm$pred[3]
  forecasts[i+1] <- forecastmt
}
cpiarima3 <- ts(forecasts, start=c(1997,04), frequency=12)
cpiarima3
# now doing a exponential smoothing model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  model <- HoltWinters(CPIs, beta=FALSE,gamma=FALSE)
  forecastm <- predict(model,n.ahead=3)
  forecastmt <-forecastm[3]
  forecasts[i+1] <- forecastmt
}
cpies3 <- ts(forecasts, start=c(1997,04), frequency=12)
cpies3

# now doing a factor model
totaldatacpi<- ts(paper.data.editedCPI,start=1997,frequency=12)
class(totaldatacpi)
dim(totaldatacpi)
a <- diff(log(totaldatacpi[,1:22]))
dim(a)
head(a)
b<- diff(totaldatacpi[,23:35])
head(b)
dtotaldata <- cbind(a,b)
```

```
dim(dtotaldata)
head(dtotaldata)
dtotaledited <- scale(dtotaldata, center=TRUE, scale=TRUE)
class(dtotaldata)
dtotaledited
class(dtotaledited)
dim(dtotaledited)
dtotaledited1 <- ts(dtotaledited[1:103,],
start=c(1997,02),frequency=12)
dim(dtotaledited1)
pca1 <- princomp(dtotaledited1, cor=TRUE)
summary(pca1)
class(pca1$scores)
dim(pca1$scores)
POETKhat(dtotaledited1)
factor1 <- ts(pca1$scores[1:103,1:1], start=c(1997,02),frequency=12)
factor1
class(factor1) # it is time series and ready to fit DI model
modelDI1 <- arima(dADcpi1,order=c(orderar[103],0,0),xreg=factor1)
forecastDI3 <- predict(modelDI1,n.ahead=3,newxreg=factor1[103])
forecastDI3
forecastDI3$pred[3]
# now rolling it up
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  dtotaleditеds <- ts(dtotaledited[1:i,],
```

```
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditEds, cor=TRUE)
  m<- POETKhat(dtotaleditEds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(CPIs,d=NA,D=NA,max.p=24,max.q=0,
  max.P=0,max.Q=0,max.d=0,max.D=0,ic=c("aic"), xreg=factors)
  forecastm <- predict(model, n.ahead=3,newxreg=factors[i])
  forecastmt <- forecastm$pred[3]
  forecasts[i+1] <- forecastmt
}
dicpi3 <- ts(forecasts, start=c(1997,04), frequency=12)
dicpi3
# now adding factors to arima


T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  dtotaleditEds <- ts(dtotaledited[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditEds, cor=TRUE)
  m<- POETKhat(dtotaleditEds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(CPIs,xreg=factors,ic=c("aic"))
  forecastm <- predict(model, n.ahead=3,newxreg=factors[i])
  forecastmt <- forecastm$pred[3]
```

```
  forecasts[i+1] <- forecastmt
}
FARIMAcpi3 <- ts(forecasts, start=c(1997,04), frequency=12)
FARIMAcpi3
# trying to do a FAVAR approach
pca1 <- princomp(dtotaledited1, cor=TRUE)\
summary(pca1)
class(pca1$scores)
dim(pca1$scores)
POETKhat(dtotaledited1)
factor1 <- ts(pca1$scores[1:103,1], start=c(1997,02),frequency=12)  #
factor1
varobject <-cbind(dADcpi1,factor1)
class(varobject)
head(varobject)
dim(varobject)
plot(varobject)
var1 <- VAR(varobject,p=1,type=c("const"),ic=c("AIC"))
predict(var1,n.ahead=1)
forecast1=forecast(var1,h=3)
forecast1$mean$dADcpi1[3]
# now doing var loop
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedS <- ts(dtotaledited[1:i,],
  start=c(1997,02),frequency=12)
```

```
  pca1 <- princomp(dtotalediteds, cor=TRUE)
  m<- POETKhat(dtotalediteds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  varcpidata <- cbind(CPIs,factors)
  model <- VAR(varcpidata,p=1,type=c("const"),ic=c("AIC"))
  forecastm <- forecast(model,h=3)
  forecastmt <- forecastm$mean$CPIs[3]
  forecasts[i+1] <- forecastmt
}
FAVARcpi3 <- ts(forecasts, start=c(1997,04), frequency=12)
FAVARcpi3

#now doing IP
tIP <- ts(adip,start=1997,frequency=12)
class(tIP)
dim(tIP)
plot(tIP, main="China's monthly seasonal adjusted industrial
production", xlab="time", ylab="billion Yuan")
adf.test(tIP)
# seems like we need to take log of first difference
lIP=log(tIP)
lIP
ldIP=diff(lIP)
class(ldIP[1:103])
IP1 <- ts(ldIP[1:103,],start=c(1997,02),frequency=12)
class(IP1)
# now doing a loop for AR
T <- 205
```

```
start <- 103
forecasts <- NA
orders <- NA
for (i in start:T){
  IPc<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  model <- ar(IPc, method="ols", aic=TRUE)
  orders[i] <- model$order
  forecastm <- predict(model, n.ahead=3)
  forecastmt <- forecastm$pred[3]
  forecasts[i+1] <- forecastmt
}
orderip3 <- ts(orders, start=c(1997,02), frequency=12)
arip3 <- ts(forecasts, start=c(1997,04), frequency=12)
orderip3
arip3

# doing arima loop
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  IPc<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  model <- auto.arima(IPc,ic=c("aic"))
  forecastm <- predict(model, n.ahead=3)
  forecastmt <- forecastm$pred[3]
  forecasts[i+1] <- forecastmt
}
iparima3 <- ts(forecasts, start=c(1997,04), frequency=12)
iparima3
```

```
#finally doing exponential smoothing for IP
T <- 205
start <- 103
forecasts <- NA
orders <- NA
for (i in start:T){
  dtc<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  model <- HoltWinters(dtc, beta=FALSE,gamma=FALSE)
  forecastm <- predict(model,n.ahead=3)
  forecastmt <-forecastm[3]
  forecasts[i+1] <- forecastmt
}
ipes3 <- ts(forecasts, start=c(1997,04), frequency=12)
ipes3

# now doing diffusion index
totaldataIP<- ts(paper.data.editedIP,start=1997,frequency=12)
class(totaldataIP)
c <- diff(log(totaldataIP[,1:21]))
dim(c)
d <- diff(totaldataIP[,22:35])
head(d)
dtotaldataIP <- cbind(c,d)
dtotaleditedIP <- scale(dtotaldataIP, center=TRUE, scale=TRUE)
dtotaleditedIP
class(dtotaleditedIP)
dim(dtotaleditedIP)
dtotaledited1IP <- ts(dtotaleditedIP[1:103,],
```

```
start=c(1997,02),frequency=12) #first within-sample period
dim(dtotaledited1IP)
pca1IP <- princomp(dtotaledited1IP, cor=TRUE)
summary(pca1IP)
class(pca1IP$scores)
dim(pca1IP$scores)
POETKhat(dtotaledited1IP)
factor1IP <- ts(pca1IP$scores[1:103,1], start=c(1997,02),frequency=12)
factor1IP
class(factor1IP)
modelDI1IP <- arima(IP1,order=c(12,0,0),xreg=factor1IP)
forecastDI1IP <- predict(modelDI1IP,n.ahead=3,newxreg=factor1IP[103])
forecastDI1IP$pred[3]
# now rolling it up
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  IPs<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  dtotaledited s <-
  ts(dtotaleditedIP[1:i,],start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotalediteds, cor=TRUE)
  m <- POETKhat(dtotalediteds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(IPs,d=NA,D=NA,max.p=24,max.q=0,
  max.P=0,max.Q=0,max.d=0,max.D=0,ic=c("aic"), xreg=factors)
  forecastm <- predict(model, n.ahead=3,newxreg=factors[i])
  forecastmt <- forecastm$pred[3]
```

```
  forecasts[i+1] <- forecastmt
}
diip3 <- ts(forecasts, start=c(1997,04), frequency=12)
diip3
# now adding factor model to arima
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  IPs<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedS <- ts(dtotaleditedIP[1:i,],
  =c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedS, cor=TRUE)
  m <- POETKhat(dtotaleditedS)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(IPs,xreg=factors,ic=c("aic"))
  forecastm <- predict(model, n.ahead=3,newxreg=factors[i])
  forecastmt <- forecastm$pred[3]
  forecasts[i+1] <- forecastmt
}
FARIMAip3 <- ts(forecasts, start=c(1997,04), frequency=12)
FARIMAip3
# now doing a FAVAR for ip
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  IPs<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
```

```
  dtotalediteds <- ts(dtotaleditedIP[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotalediteds, cor=TRUE)
  m <- POETKhat(dtotalediteds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  varipdata <- cbind(IPs,factors)
  model <- VAR(varipdata,p=1,type=c("const"),ic=c("AIC"))
  forecastm <- forecast(model,h=3)
  forecastmt <- forecastm$mean$IPs[3]
  forecasts[i+1] <- forecastmt
}
FAVARIP3 <- ts(forecasts, start=c(1997,04), frequency=12)
FAVARIP3
# now doing a electricity production
adPE <- ts(adpe,start=1997,frequency=12)
dim(adPE)
plot(adPE)
adf.test(adPE)
dladPE <- diff(log(adPE)) # take first difference
plot(dladPE, main="percentage change of prodcution of electricity",
xlab="time", ylab="percentage", col="blue")
plot(adPE, main="volume of production of electricity", xlab="time",
ylab="volume")
# now fit to ar model first
dladPE1 <- ts(dladPE[1:103],start=c(1997,02),frequency=12)
model1 <- ar(dladPE,method="ols",aic=TRUE)
model1
predict(model1,n.ahead=1)
```

```
forecast(model1,h=1)

forecast(model1,h=3)

predict(model1,n.ahead=3)$pred[3]


# now doing a loop

T <- 205

start <- 103

forecasts <- NA

orders <- NA

for (i in start:T){

  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)

  model <- ar(PEs, method="ols", aic=TRUE)

  orders[i] <- model$order

  forecastm <- predict(model, n.ahead=3)

  forecastmt <- forecastm$pred[3]

  forecasts[i+1] <- forecastmt

}

orderpe3 <- ts(orders, start=c(1997,02), frequency=12)

pear3 <- ts(forecasts, start=c(1997,04), frequency=12)

pear3 #this is AR forecast for PPI

orderpe3 # this is AR order that we need to use in later factor model

# now doing arima for AR

T <- 205

start <- 103

forecasts <- NA

for (i in start:T){

  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)

  model <- auto.arima(PEs,ic=c("aic"))

  forecastm <- predict(model, n.ahead=3)
```

```
  forecastmt <- forecastm$pred[3]
  forecasts[i+1] <- forecastmt
}
pearima3 <- ts(forecasts, start=c(1997,04), frequency=12)
pearima3
# now doing a exponential smoothing model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  model <- HoltWinters(PEs, beta=FALSE,gamma=FALSE)
  forecastm <- predict(model,n.ahead=3)
  forecastmt <-forecastm[3]
  forecasts[i+1] <- forecastmt
}
pees3 <- ts(forecasts, start=c(1997,04), frequency=12)
pees3

# now doing a factor model
totaldatape<- ts(paper.data.editedPE,start=1997,frequency=12)
class(totaldatape)
e <- diff(log(totaldatacpi[,1:21]))
dim(e)
head(e)
f<- diff(totaldatacpi[,22:35])
head(f)
dtotaldatape <- cbind(e,f)
dtotaleditedpe <- scale(dtotaldatape, center=TRUE, scale=TRUE)
```

```
dtotaleditedpe

class(dtotaleditedpe)

dim(dtotaleditedpe)

dtotaleditedpe1 <- ts(dtotaleditedpe[1:103,],

start=c(1997,02),frequency=12)

dim(dtotaleditedpe1)

pca1 <- princomp(dtotaleditedpe1, cor=TRUE)

summary(pca1)

class(pca1$scores)

dim(pca1$scores)

POETKhat(dtotaleditedpe1)

factor1 <- ts(pca1$scores[1:103,1], start=c(1997,02),frequency=12)  #

factor1

class(factor1) # it is time series and ready to fit DI model

modelDI1 <- arima(dladPE1,order=c(orderpe[103],0,0),xreg=factor1)

forecastDI1 <- predict(modelDI1,n.ahead=1,newxreg=factor1[103])

forecastDI1

# now rolling it up

T <- 205

start <- 103

forecasts <- NA

for (i in start:T){

  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)

  dtotaleditedpes <- ts(dtotaleditedpe[1:i,],

  start=c(1997,02),frequency=12)

  pca1 <- princomp(dtotaleditedpes, cor=TRUE)

  m<- POETKhat(dtotaleditedpes)

  factors <- ts(pca1$scores[1:i,1:m$K1BN],

  start=c(1997,02),frequency=12)
```

```
  model <- auto.arima(PEs,d=NA,D=NA,max.p=24,max.q=0,
  max.P=0,max.Q=0,max.d=0,max.D=0,ic=c("aic"), xreg=factors)
  forecastm <- predict(model, n.ahead=3,newxreg=factors[i])
  forecastmt <- forecastm$pred[3]
  forecasts[i+1] <- forecastmt
}
dipe3 <- ts(forecasts, start=c(1997,04), frequency=12)
dipe3
# now adding the factor to arima model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedpes <- ts(dtotaleditedpe[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedpes, cor=TRUE)
  m<- POETKhat(dtotaleditedpes)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(PEs,xreg=factors,ic=c("aic"))
  forecastm <- predict(model, n.ahead=3,newxreg=factors[i])
  forecastmt <- forecastm$pred[3]
  forecasts[i+1] <- forecastmt
}
FARIMApe3 <- ts(forecasts, start=c(1997,04), frequency=12)
FARIMApe3
# now doing FAVAR for PE
T <- 205
```

```
start <- 103
forecasts <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedpes <- ts(dtotaleditedpe[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedpes, cor=TRUE)
  m<- POETKhat(dtotaleditedpes)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  varipdata <- cbind(PEs,factors)
  model <- VAR(varipdata,p=1,type=c("const"),ic=c("AIC"))
  forecastm <- forecast(model,h=3)
  forecastmt <- forecastm$mean$PEs[3]
  forecasts[i+1] <- forecastmt
}
FAVARpe3 <- ts(forecasts, start=c(1997,04), frequency=12)
FAVARpe3

# now doing producer price index
adPPI <- ts(adppi,start=1997,frequency=12)
dim(adPPI )
plot(adPPI )
adf.test(adPPI )
dadPPI  <- diff(adPPI ) # take first difference
plot(dadPPI , main="first difference of producer price index",
xlab="time", ylab="%change", col="blue")
plot(adPPI, main="monthly producer price index", xlab="time",
ylab="%")
```

```
# now fit to ar model first
dadPPI1 <- ts(dadPPI[1:103],start=c(1997,02),frequency=12)
model1 <- ar(dadPPI1,method="ols",aic=TRUE)
model1
predict(model1,n.ahead=3)
# now doing a loop
T <- 205
start <- 103
forecasts <- NA
orders <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  model <- ar(PPIs, method="ols", aic=TRUE)
  orders[i] <- model$order
  forecastm <- predict(model, n.ahead=3)
  forecastmt <- forecastm$pred[3]
  forecasts[i+1] <- forecastmt
}
orderppi3 <- ts(orders, start=c(1997,02), frequency=12)
ppiar3 <- ts(forecasts, start=c(1997,04), frequency=12)
ppiar3 #this is AR forecast for CPI
orderppi3
# now doing arima for AR
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  model <- auto.arima(PPIs,ic=c("aic"))
```

```
  forecastm <- predict(model, n.ahead=3)
  forecastmt <- forecastm$pred[3]
  forecasts[i+1] <- forecastmt
}
ppiarima3 <- ts(forecasts, start=c(1997,04), frequency=12)
ppiarima3
# now doing a exponential smoothing model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  model <- HoltWinters(PPIs, beta=FALSE,gamma=FALSE)
  forecastm <- predict(model,n.ahead=3)
  forecastmt <-forecastm[3]
  forecasts[i+1] <- forecastmt
}
ppies3 <- ts(forecasts, start=c(1997,04), frequency=12)
ppies3

# now doing a factor model
totaldatappi<- ts(paper.data.editedPPI,start=1997,frequency=12)
class(totaldatappi)
g <- diff(log(totaldatacpi[,1:22]))
dim(g)
head(g)
h<- diff(totaldatacpi[,23:35])
head(h)
dtotaldatappi <- cbind(g,h)
```

```
dim(totaldatappi)
dtotaldatappied <- scale(dtotaldatappi, center=TRUE, scale=TRUE)
class(dtotaldatappied)
dim(dtotaldatappied)
dtotaldatappied1 <- ts(dtotaldatappied[1:103,],
start=c(1997,02),frequency=12) #first within-sample period
dim(dtotaldatappied1)
pca1 <- princomp(dtotaldatappied1, cor=TRUE)
summary(pca1)
class(pca1$scores)
dim(pca1$scores)
POETKhat(dtotaldatappied1)
factor1 <- ts(pca1$scores[1:103,1], start=c(1997,02),frequency=12)  #
factor1
class(factor1)
modelDI1 <- arima(dadPPI1,order=c(orderpe[103],0,0),xreg=factor1)
forecastDI1 <- predict(modelDI1,n.ahead=1,newxreg=factor1[103])
forecastDI1
# now rolling it up
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  dtotaldatappieds <- ts(dtotaldatappied[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaldatappieds, cor=TRUE)
  m<- POETKhat(dtotaldatappieds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
```

```
  start=c(1997,02),frequency=12)
  model <- auto.arima(PPIs,d=NA,D=NA,max.p=24,max.q=0,
  max.P=0,max.Q=0,max.d=0,max.D=0,ic=c("aic"), xreg=factors)
  forecastm <- predict(model, n.ahead=3,newxreg=factors[i])
  forecastmt <- forecastm$pred[3]
  forecasts[i+1] <- forecastmt
}
dippi3 <- ts(forecasts, start=c(1997,04), frequency=12)
dippi3
# now adding the factors to arima model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  dtotaldatappieds <- ts(dtotaldatappied[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaldatappieds, cor=TRUE)
  m<- POETKhat(dtotaldatappieds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(PPIs,xreg=factors,ic=c("aic"))
  forecastm <- predict(model, n.ahead=3,newxreg=factors[i])
  forecastmt <- forecastm$pred[3]
  forecasts[i+1] <- forecastmt
}
FARIMAppi3 <- ts(forecasts, start=c(1997,04), frequency=12)
FARIMAppi3
# now doing FAVAR for PPI
```

```
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  dtotaldatappieds <- ts(dtotaldatappied[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaldatappieds, cor=TRUE)
  m<- POETKhat(dtotaldatappieds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  varipdata <- cbind(PPIs,factors)
  model <- VAR(varipdata,p=1,type=c("const"),ic=c("AIC"))
  forecastm <- forecast(model,h=3)
  forecastmt <- forecastm$mean$PPIs[3]
  forecasts[i+1] <- forecastmt
}
FAVARppi3 <- ts(forecasts, start=c(1997,04), frequency=12)
FAVARppi3
# now writing user defined functions to compute MSE and AME
mse <- function(x,y,n){
  sum((x-y)^2)/n
}
n <- 101 # this is total number of out-of-samle period

cpiAR3 <- mse(dADcpi[106:206],cpiar3[106:206],n)
cpiARIMA3 <- mse(dADcpi[106:206],cpiarima3[106:206],n)
cpiES3 <- mse(dADcpi[106:206],cpies3[106:206],n)
cpiDI3 <- mse(dADcpi[106:206],dicpi3[106:206],n)
```

```
cpiFARIMA3 <- mse(dADcpi[106:206],FARIMAcpi3[106:206],n)
cpiFAVAR3 <- mse(dADcpi[106:206],FAVARcpi3[106:206],n)
CPIforecastMSE3 <-cbind(cpiAR3,cpiARIMA3,
cpiES3,cpiDI3,cpiFARIMA3,cpiFAVAR3)
CPIforecastMSE3
tsCPI3 <- ts(dADcpi[106:206],start=c(2005,11),frequency=12)
tscpiar3 <-ts(cpiar3[106:206],start=c(2005,11),frequency=12)
tscpiarima3 <-ts(cpiarima3[106:206],start=c(2005,11),frequency=12)
tscpies3 <-ts(cpies3[106:206],start=c(2005,11),frequency=12)
tscpidi3 <-ts(dicpi3[106:206],start=c(2005,11),frequency=12)
tscpiFARIMA3 <-ts(FARIMAcpi3[106:206],start=c(2005,11),frequency=12)
tscpiFAVAR3 <-ts(FAVARcpi3[106:206],start=c(2005,11),frequency=12)
ts.plot(tsCPI3,tscpiar3,
        gpars=list(main="plot of CPI AR forecast and actual
        value",xlab="time",ylab="difference of CPI",lty=c(1:2)))
ts.plot(tsCPI3,tscpiarima3,
        gpars=list(main="plot of CPI ARIMA forecast and actual
        value",xlab="time",ylab="difference of CPI",lty=c(1:2)))
ts.plot(tsCPI3,tscpies3,
        gpars=list(main="plot of CPI EM forecast and actual
        value",xlab="time",ylab="difference of CPI",lty=c(1:2)))
ts.plot(tsCPI3,tscpidi3,
        gpars=list(main="plot of CPI DI forecast and actual
        value",xlab="time",ylab="difference of CPI",lty=c(1:2)))
ts.plot(tsCPI3,tscpiFARIMA3,
        gpars=list(main="plot of CPI FARIMA forecast and actual
        value",xlab="time",ylab="difference of CPI",lty=c(1:2)))
ts.plot(tsCPI3,tscpiFAVAR3,
       gpars=list(main="plot of CPI FAVAR forecast and actual
```

```
       value",xlab="time",ylab="difference of CPI",lty=c(1:2)))
# now for IP
ipAR3 <- mse(ldIP[106:206],arip3[106:206],n)
ipARIMA3 <- mse(ldIP[106:206],iparima3[106:206],n)
ipES3 <- mse(ldIP[106:206],ipes3[106:206],n)
ipDI3 <- mse(ldIP[106:206],diip3[106:206],n)
ipFARIMA3 <- mse(ldIP[106:206],FARIMAip3[106:206],n)
ipFAVAR3 <-mse(ldIP[106:206],FAVARIP3[106:206],n)
IPforecastMSE3 <-cbind(ipAR3,ipARIMA3,ipES3,ipDI3,ipFARIMA3,ipFAVAR3)
IPforecastMSE3
IP3 <- ts(ldIP[106:206],start=c(2005,11),frequency=12)
tsipar3 <-ts(arip3[106:206],start=c(2005,11),frequency=12)
tsiparima3 <-ts(iparima3[106:206],start=c(2005,11),frequency=12)
tsipes3 <-ts(ipes3[106:206],start=c(2005,11),frequency=12)
tsipdi3 <-ts(diip3[106:206],start=c(2005,11),frequency=12)
tsipFARIMA3 <-ts(FARIMAip3[106:206],start=c(2005,11),frequency=12)
tsipFAVAR3 <-ts(FAVARIP3[106:206],start=c(2005,11),frequency=12)
ts.plot(tsIP3,tsipar3,
        gpars=list(main="plot of IP AR forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsIP3,tsiparima3,
        gpars=list(main="plot of IP ARIMA forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsIP3,tsipes3,
        gpars=list(main="plot of IP EM forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsIP3,tsipdi3,
        gpars=list(main="plot of IP DI forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
```

```
ts.plot(tsIP3,tsipFARIMA3,
        gpars=list(main="plot of IP FARIMA forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsIP3,tsipFAVAR3,
        gpars=list(main="plot of IP FAVAR forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
# now for Production of electricity
peAR3 <- mse(dladPE[106:206],pear3[106:206],n)
peARIMA3 <- mse(dladPE[106:206],pearima3[106:206],n)
peES3 <- mse(dladPE[106:206],pees3[106:206],n)
peDI3 <- mse(dladPE[106:206],dipe3[106:206],n)
peFARIMA3 <- mse(dladPE[106:206],FARIMApe3[106:206],n)
peFAVAR3 <- mse(dladPE[106:206],FAVARpe3[106:206],n)
PEforecastMSE3<-cbind(peAR3,peARIMA3,peES3,peDI3,peFARIMA3,peFAVAR3)
PEforecastMSE3

PE3 <- ts(dladPE[106:206],start=c(2005,11),frequency=12)
tspear3 <-ts(pear3[106:206],start=c(2005,11),frequency=12)
tspearima3 <-ts(pearima3[106:206],start=c(2005,11),frequency=12)
tspees3 <-ts(pees3[106:206],start=c(2005,11),frequency=12)
tspedi3 <-ts(dipe3[106:206],start=c(2005,11),frequency=12)
tspeFARIMA3 <-ts(FARIMApe3[106:206],start=c(2005,11),frequency=12)
tspeFAVAR3 <-ts(FAVARpe3[106:206],start=c(2005,11),frequency=12)
ts.plot(tsPE3,tspear3,
        gpars=list(main="plot of PE AR forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPE3,tspearima3,
        gpars=list(main="plot of PE ARIMA forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
```

```
ts.plot(tsPE3,tspees3,
        gpars=list(main="plot of PE EM forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPE3,tspedi3,
        gpars=list(main="plot of PE DI forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPE3,tspeFARIMA3,
        gpars=list(main="plot of PE FARIMA forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPE3,tspeFAVAR3,
        gpars=list(main="plot of PE FAVAR forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))

#now for PPI
ppiAR3 <- mse(dadPPI[106:206],ppiar3[106:206],n)
ppiARIMA3 <- mse(dadPPI[106:206],ppiarima3[106:206],n)
ppiES3 <- mse(dadPPI[106:206],ppies3[106:206],n)
ppiDI3 <- mse(dadPPI[106:206],dippi3[106:206],n)
ppiFARIMA3 <- mse(dadPPI[106:206],FARIMAppi3[106:206],n)
ppiFAVAR3 <- mse(dadPPI[106:206],FAVARppi3[106:206],n)

tsPPI3 <- ts(dadPPI[106:206],start=c(2005,11),frequency=12)
tsppiar3 <-ts(ppiar3[106:206],start=c(2005,11),frequency=12)
tsppiarima3 <-ts(ppiarima3[106:206],start=c(2005,11),frequency=12)
tsppies3 <-ts(ppies3[106:206],start=c(2005,11),frequency=12)
tsppidi3 <-ts(dippi3[106:206],start=c(2005,11),frequency=12)
tsppiFARIMA3 <-ts(FARIMAppi3[106:206],start=c(2005,11),frequency=12)
tsppiFAVAR3 <-ts(FAVARppi3[106:206],start=c(2005,11),frequency=12)
ts.plot(tsPPI3,tsppiar3,
```

```
        gpars=list(main="plot of PPI AR forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPPI3,tsppiarima3,
        gpars=list(main="plot of PPI ARIMA forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPPI3,tsppies3,
        gpars=list(main="plot of PPI EM forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPPI3,tsppidi3,
        gpars=list(main="plot of PPI DI forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPPI3,tsppiFARIMA3,
        gpars=list(main="plot of PPI FARIMA forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPPI3,tsppiFAVAR3,
        gpars=list(main="plot of PPI FAVAR forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
PPIforecastMSE3 <- cbind(ppiAR3,ppiARIMA3,
ppiES3,ppiDI3,ppiFARIMA3,ppiFAVAR3)
CPIforecastMSE3
IPforecastMSE3
PEforecastMSE3
PPIforecastMSE3
totalforecast3 <-rbind(CPIforecastMSE3,
IPforecastMSE3,PEforecastMSE3,PPIforecastMSE3)
totalforecast3
colnames(totalforecast3) <-
c("AR","ARIMA","ES","DI","FARIMA","FAVAR")
rownames(totalforecast3) <- c("CPI","IP","PE","PPI")
```

```
totalMSE3 <-t(totalforecast3)
totalMSE3

# now doing relatively MSE and MAS
RCPImse3 <- totalMSE3[,1]/totalMSE3[1,1]
RIPmse3 <- totalMSE3[,2]/totalMSE3[1,2]
RPEmse3 <- totalMSE3[,3]/totalMSE3[1,3]
RPPImse3 <-totalMSE3[,4]/totalMSE3[1,4]
RCPImse3
RIPmse3
RPEmse3
RPPImse3
rbind(RCPImse3,RIPmse3,RPEmse3,RPPImse3)
RtotalMSE3 <- t(rbind(RCPImse3,RIPmse3,RPEmse3,RPPImse3))
colnames(RtotalMSE3) <- c("CPI","IP","PE","PPI")
RtotalMSE3
totalMSFE3 <- xtable(RtotalMSE3)
print.xtable(totalMAFE3,type="latex", file="",floating=FALSE,table.placement="H")
print.xtable(totalMSFE3,type="latex", file="",floating=TRUE,table.placement="H")

#six-month-ahead forecasts
ADcpi<- ts(adcpi,start=1997,frequency=12)
dim(ADcpi)
plot(ADcpi, main="month-on-month CPI", xlab="time", ylab="%")
adf.test(ADcpi)
China'scpi <- ts(China's.CPI.MoM, start=1997,frequency=12)
plot(China'scpi)
acf(China'scpi,lag.max=24)
acf(ADcpi, lag.max=24)
```

```
dADcpi <- diff(ADcpi) # take first difference
plot(dADcpi, main="seasonally adjusted China's monthly inflation
rate", xlab="time", ylab="percentage", col="blue")
# now fit to ar model first
dADcpi1 <- ts(dADcpi[1:103],start=c(1997,02),frequency=12)
model1 <- ar(dADcpi1,method="ols",aic=TRUE)
model1
predict(model1,n.ahead=1)
predict(model1,n.ahead=6)$pred[6]
forecast(model1,h=6)

dADcpi2 <- ts(dADcpi[1:104],start=c(1997,02),frequency=12)
model2<- ar(dADcpi2,method="ols",aic=TRUE)
forecast(model2,h=6)
predict(model2,n.ahead=6)
predict(model2,n.ahead=6)$pred[6]

# now doing a loop
T <- 205
start <- 103
forecasts <- NA
orders <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  model <- ar(CPIs, method="ols", aic=TRUE)
  orders[i] <- model$order
  forecastm <- predict(model, n.ahead=6)
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
```

```
}
orderar6 <- ts(orders, start=c(1997,02), frequency=12)
cpiar6 <- ts(forecasts, start=c(1997,07), frequency=12)
cpiar6 #this is AR forecast for CPI
orderar6 # this is AR order that we need to use in later factor model
# now doing arima for AR
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  model <- auto.arima(CPIs,ic=c("aic"))
  forecastm <- predict(model, n.ahead=6)
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
cpiarima6 <- ts(forecasts, start=c(1997,07), frequency=12)
cpiarima6
# now doing a exponential smoothing model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  model <- HoltWinters(CPIs, beta=FALSE,gamma=FALSE)
  forecastm <- predict(model,n.ahead=6)
  forecastmt <-forecastm[6]
  forecasts[i+1] <- forecastmt
}
```

```
cpies6 <- ts(forecasts, start=c(1997,07), frequency=12)
cpies6


# now doing a factor model
totaldatacpi<- ts(paper.data.editedCPI,start=1997,frequency=12)
class(totaldatacpi)
dim(totaldatacpi)
a <- diff(log(totaldatacpi[,1:22]))
dim(a)
head(a)
b<- diff(totaldatacpi[,23:35])
head(b)
dtotaldata <- cbind(a,b)
dim(dtotaldata)
head(dtotaldata)
dtotaledited <- scale(dtotaldata, center=TRUE, scale=TRUE)
class(dtotaldata)
dtotaledited
class(dtotaledited)
dim(dtotaledited)
dtotaledited1 <- ts(dtotaledited[1:103,],
start=c(1997,02),frequency=12)
dim(dtotaledited1)
pca1 <- princomp(dtotaledited1, cor=TRUE)
summary(pca1)
class(pca1$scores)
dim(pca1$scores)
POETKhat(dtotaledited1)
factor1 <- ts(pca1$scores[1:103,1:1], start=c(1997,02),frequency=12)
```

```
factor1

class(factor1)

modelDI1 <- arima(dADcpi1,order=c(orderar6[103],0,0),xreg=factor1)

forecastDI6 <- predict(modelDI1,n.ahead=6,newxreg=factor1[103])

forecastDI6

forecastDI6$pred[6]

# now rolling it up

T <- 205

start <- 103

forecasts <- NA

for (i in start:T){

  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)

  dtotaleditedS <- ts(dtotaledited[1:i,],

  start=c(1997,02),frequency=12)

  pca1 <- princomp(dtotaleditedS, cor=TRUE)

  m<- POETKhat(dtotaleditedS)

  factors <- ts(pca1$scores[1:i,1:m$K1BN],

  start=c(1997,02),frequency=12)

  model <- auto.arima(CPIs,d=NA,D=NA,max.p=24,max.q=0,max.P=0,

  max.Q=0,max.d=0,max.D=0,ic=c("aic"), xreg=factors)

  forecastm <- predict(model, n.ahead=6,newxreg=factors[i])

  forecastmt <- forecastm$pred[6]

  forecasts[i+1] <- forecastmt

}

dicpi6 <- ts(forecasts, start=c(1997,07), frequency=12)

dicpi6

# now adding factors to arima


T <- 205
```

```
start <- 103
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  dtotaledited s <- ts(dtotaledited[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaledited s, cor=TRUE)
  m<- POETKhat(dtotaledited s)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(CPIs,xreg=factors,ic=c("aic"))
  forecastm <- predict(model, n.ahead=6,newxreg=factors[i])
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
FARIMAcpi6 <- ts(forecasts, start=c(1997,07), frequency=12)
FARIMAcpi6
# trying to do a FAVAR approach
pca1 <- princomp(dtotaledited1, cor=TRUE)
summary(pca1)
class(pca1$scores)
dim(pca1$scores)
POETKhat(dtotaledited1)
factor1 <- ts(pca1$scores[1:103,1], start=c(1997,02),frequency=12)  #
factor1
varobject <-cbind(dADcpi1,factor1)
class(varobject)
head(varobject)
dim(varobject)
```

```
plot(varobject)
var1 <- VAR(varobject,p=1,type=c("const"),ic=c("AIC"))
predict(var1,n.ahead=6)
forecast1=forecast(var1,h=6)
forecast1$mean$dADcpi1[6]

# now doing var loop
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  CPIs<- ts(dADcpi[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedS <- ts(dtotaledited[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedS, cor=TRUE)
  m<- POETKhat(dtotaleditedS)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  varcpidata <- cbind(CPIs,factors)
  model <- VAR(varcpidata,p=1,type=c("const"),ic=c("AIC"))
  forecastm <- forecast(model,h=6)
  forecastmt <- forecastm$mean$CPIs[6]
  forecasts[i+1] <- forecastmt
}
FAVARcpi6 <- ts(forecasts, start=c(1997,07), frequency=12)
FAVARcpi6

#now doing IP
tIP <- ts(adip,start=1997,frequency=12)
```

```
class(tIP)
dim(tIP)
plot(tIP, main="China's monthly seasonal adjusted industrial
production", xlab="time", ylab="billion Yuan")
adf.test(tIP)
# seems like we need to take log of first difference
lIP=log(tIP)
lIP
ldIP=diff(lIP)
class(ldIP[1:103])
IP1 <- ts(ldIP[1:103,],start=c(1997,02),frequency=12)
class(IP1)
# now doing a loop for AR
T <- 205
start <- 103
forecasts <- NA
orders <- NA
for (i in start:T){
  IPc<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  model <- ar(IPc, method="ols", aic=TRUE)
  orders[i] <- model$order
  forecastm <- predict(model, n.ahead=6)
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
orderip6 <- ts(orders, start=c(1997,02), frequency=12)
arip6 <- ts(forecasts, start=c(1997,07), frequency=12)
orderip6
arip6
```

```
# doing arima loop
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  IPc<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  model <- auto.arima(IPc,ic=c("aic"))
  forecastm <- predict(model, n.ahead=6)
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
iparima6 <- ts(forecasts, start=c(1997,07), frequency=12)
iparima6

#finally doing exponential smoothing for IP
T <- 205
start <- 103
forecasts <- NA
orders <- NA
for (i in start:T){
  dtc<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  model <- HoltWinters(dtc, beta=FALSE,gamma=FALSE)
  forecastm <- predict(model,n.ahead=6)
  forecastmt <-forecastm[6]
  forecasts[i+1] <- forecastmt
}
ipes6 <- ts(forecasts, start=c(1997,07), frequency=12)
ipes6
```

```
# now doing diffusion index
totaldataIP<- ts(paper.data.editedIP,start=1997,frequency=12)
class(totaldataIP)
c <- diff(log(totaldataIP[,1:21]))
dim(c)
d <- diff(totaldataIP[,22:35])
head(d)
dtotaldataIP <- cbind(c,d)
dtotaleditedIP <- scale(dtotaldataIP, center=TRUE, scale=TRUE)
dtotaleditedIP
class(dtotaleditedIP)
dim(dtotaleditedIP)
dtotaledited1IP <-
ts(dtotaleditedIP[1:103,],start=c(1997,02),frequency=12)
dim(dtotaledited1IP)
pca1IP <- princomp(dtotaledited1IP, cor=TRUE)
summary(pca1IP)
class(pca1IP$scores)
dim(pca1IP$scores)
POETKhat(dtotaledited1IP)
factor1IP <- ts(pca1IP$scores[1:103,1],
start=c(1997,02),frequency=12)
factor1IP
class(factor1IP)
modelDI1IP <- arima(IP1,order=c(12,0,0),xreg=factor1IP)
forecastDI1IP <- predict(modelDI1IP,n.ahead=6,newxreg=factor1IP[103])
forecastDI1IP$pred
forecastDI1IP$pred[6]
```

```
# now rolling it up
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  IPs<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedS <- ts(dtotaleditedIP[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedS, cor=TRUE)
  m <- POETKhat(dtotaleditedS)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(IPs,d=NA,D=NA,max.p=24,max.q=0,max.P=0,
  max.Q=0,max.d=0,max.D=0,ic=c("aic"), xreg=factors)
  forecastm <- predict(model, n.ahead=6,newxreg=factors[i])
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
diip6 <- ts(forecasts, start=c(1997,07), frequency=12)
diip6
# now adding factor model to arima
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  IPs<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedS <- ts(dtotaleditedIP[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedS, cor=TRUE)
```

```
  m <- POETKhat(dtotaleditedds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(IPs,xreg=factors)
  forecastm <- predict(model, n.ahead=6,newxreg=factors[i])
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
FARIMAip6 <- ts(forecasts, start=c(1997,07), frequency=12)
FARIMAip6
# now doing a FAVAR for ip
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  IPs<- ts(ldIP[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedds <- ts(dtotaleditedIP[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedds, cor=TRUE)
  m <- POETKhat(dtotaleditedds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  varipdata <- cbind(IPs,factors)
  model <- VAR(varipdata,p=1,type=c("const"),ic=c("AIC"))
  forecastm <- forecast(model,h=6)
  forecastmt <- forecastm$mean$IPs[6]
  forecasts[i+1] <- forecastmt
}
FAVARIP6 <- ts(forecasts, start=c(1997,07), frequency=12)
```

```
FAVARIP6
# now doing a electricity production
adPE <- ts(adpe,start=1997,frequency=12)
dim(adPE)
plot(adPE)
adf.test(adPE)
dladPE <- diff(log(adPE)) # take first difference
plot(dladPE, main="percentage change of prodcution of electricity",
xlab="time", ylab="percentage", col="blue")
plot(adPE, main="volume of production of electricity", xlab="time",
ylab="volume")
# now fit to ar model first
dladPE1 <- ts(dladPE[1:103],start=c(1997,02),frequency=12)
model1 <- ar(dladPE,method="ols",aic=TRUE)
model1
predict(model1,n.ahead=1)
forecast(model1,h=1)
forecast(model1,h=6)
predict(model1,n.ahead=6)$pred[6]

# now doing a loop
T <- 205
start <- 103
forecasts <- NA
orders <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  model <- ar(PEs, method="ols", aic=TRUE)
  orders[i] <- model$order
```

```
  forecastm <- predict(model, n.ahead=6)
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
orderpe6 <- ts(orders, start=c(1997,02), frequency=12)
pear6 <- ts(forecasts, start=c(1997,07), frequency=12)
pear6 #this is AR forecast for CPI
orderpe6 # this is AR order that we need to use in later factor model
# now doing arima for AR
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  model <- auto.arima(PEs,ic=c("aic"))
  forecastm <- predict(model, n.ahead=6)
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
pearima6 <- ts(forecasts, start=c(1997,07), frequency=12)
pearima6
# now doing a exponential smoothing model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  model <- HoltWinters(PEs, beta=FALSE,gamma=FALSE)
  forecastm <- predict(model,n.ahead=6)
```

```
  forecastmt <-forecastm[6]
  forecasts[i+1] <- forecastmt
}
pees6 <- ts(forecasts, start=c(1997,07), frequency=12)
pees6

# now doing a factor model
totaldatape<- ts(paper.data.editedPE,start=1997,frequency=12)
class(totaldatape)
e <- diff(log(totaldatacpi[,1:21]))
dim(e)
head(e)
f<- diff(totaldatacpi[,22:35])
head(f)
dtotaldatape <- cbind(e,f)
dtotaleditedpe <- scale(dtotaldatape, center=TRUE, scale=TRUE)
dtotaleditedpe
class(dtotaleditedpe)
dim(dtotaleditedpe)
dtotaleditedpe1 <- ts(dtotaleditedpe[1:103,],
start=c(1997,02),frequency=12) #first within-sample period
dim(dtotaleditedpe1)
pca1 <- princomp(dtotaleditedpe1, cor=TRUE)
summary(pca1)
class(pca1$scores)
dim(pca1$scores)
POETKhat(dtotaleditedpe1)
factor1 <- ts(pca1$scores[1:103,1], start=c(1997,02),frequency=12)  #
factor1
```

```
class(factor1)
modelDI1 <- arima(dladPE1,order=c(orderpe[103],0,0),xreg=factor1)
forecastDI1 <- predict(modelDI1,n.ahead=6,newxreg=factor1[103])
forecastDI1$pred[6]
# now rolling it up
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedpes <- ts(dtotaleditedpe[1:i,],s
  tart=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedpes, cor=TRUE)
  m<- POETKhat(dtotaleditedpes)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(PEs,d=NA,D=NA,max.p=24,max.q=0,max.P=0,
  max.Q=0,max.d=0,max.D=0,ic=c("aic"), xreg=factors)
  forecastm <- predict(model, n.ahead=6,newxreg=factors[i])
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
dipe6 <- ts(forecasts, start=c(1997,07), frequency=12)
dipe6
# now adding the factor to arima model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
```

```
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedpes <- ts(dtotaleditedpe[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedpes, cor=TRUE)
  m<- POETKhat(dtotaleditedpes)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(PEs,xreg=factors,ic=c("aic"))
  forecastm <- predict(model, n.ahead=6,newxreg=factors[i])
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
FARIMApe6 <- ts(forecasts, start=c(1997,07), frequency=12)
FARIMApe6
# now doing FAVAR for PE
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PEs<- ts(dladPE[1:i], start=c(1997, 02), frequency=12)
  dtotaleditedpes <- ts(dtotaleditedpe[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaleditedpes, cor=TRUE)
  m<- POETKhat(dtotaleditedpes)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  varipdata <- cbind(PEs,factors)
  model <- VAR(varipdata,p=1,type=c("const"),ic=c("AIC"))
  forecastm <- forecast(model,h=6)
```

```
  forecastmt <- forecastm$mean$PEs[6]
  forecasts[i+1] <- forecastmt
}
FAVARpe6 <- ts(forecasts, start=c(1997,07), frequency=12)
FAVARpe6

# now doing producer price index
adPPI <- ts(adppi,start=1997,frequency=12)
dim(adPPI )
plot(adPPI )
adf.test(adPPI )
dadPPI  <- diff(adPPI ) # take first difference
plot(dadPPI , main="first difference of producer price index"
, xlab="time", ylab="%change", col="blue")
plot(adPPI, main="monthly producer price index", xlab="time",
ylab="%")
# now fit to ar model first
dadPPI1 <- ts(dadPPI[1:103],start=c(1997,02),frequency=12)
model1 <- ar(dadPPI1,method="ml",aic=TRUE)
model1
predict(model1,n.ahead=3)
# now doing a loop
T <- 205
start <- 103
forecasts <- NA
orders <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  model <- ar(PPIs, method="ols", aic=TRUE)
```

```
  orders[i] <- model$order
  forecastm <- predict(model, n.ahead=6)
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
orderppi6 <- ts(orders, start=c(1997,02), frequency=12)
ppiar6 <- ts(forecasts, start=c(1997,07), frequency=12)
ppiar6 #this is AR forecast for CPI
orderppi6
# now doing arima for AR
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  model <- auto.arima(PPIs,ic=c("aic"))
  forecastm <- predict(model, n.ahead=6)
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
ppiarima6 <- ts(forecasts, start=c(1997,07), frequency=12)
ppiarima6
# now doing a exponential smoothing model
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  model <- HoltWinters(PPIs, beta=FALSE,gamma=FALSE)
```

```
  forecastm <- predict(model,n.ahead=6)
  forecastmt <-forecastm[6]
  forecasts[i+1] <- forecastmt
}
ppies6 <- ts(forecasts, start=c(1997,07), frequency=12)
ppies6

# now doing a factor model
totaldatappi<- ts(paper.data.editedPPI,start=1997,frequency=12)
class(totaldatappi)
g <- diff(log(totaldatacpi[,1:22]))
dim(g)
head(g)
h<- diff(totaldatacpi[,23:35])
head(h)
dtotaldatappi <- cbind(g,h)
dim(totaldatappi)
dtotaldatappied <- scale(dtotaldatappi, center=TRUE, scale=TRUE)
class(dtotaldatappied)
dim(dtotaldatappied)
dtotaldatappied1 <- ts(dtotaldatappied[1:103,],
start=c(1997,02),frequency=12)
dim(dtotaldatappied1)
pca1 <- princomp(dtotaldatappied1, cor=TRUE)
summary(pca1)
class(pca1$scores)
dim(pca1$scores)
POETKhat(dtotaldatappied1)
factor1 <- ts(pca1$scores[1:103,1], start=c(1997,02),frequency=12)  #
```

```
factor1
class(factor1)
modelDI1 <- arima(dadPPI1,order=c(orderpe[103],0,0),xreg=factor1)
forecastDI1 <- predict(modelDI1,n.ahead=1,newxreg=factor1[103])
forecastDI1
# now rolling it up
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  dtotaldatappieds <-
  ts(dtotaldatappied[1:i,],start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaldatappieds, cor=TRUE)
  m<- POETKhat(dtotaldatappieds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(PPIs,d=NA,D=NA,max.p=24,max.q=0,max.P=0,
  max.Q=0,max.d=0,max.D=0,ic=c("aic"), xreg=factors)
  forecastm <- predict(model, n.ahead=6,newxreg=factors[i])
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
dippi6 <- ts(forecasts, start=c(1997,07), frequency=12)
dippi6
# now adding the factors to arima model
T <- 205
start <- 103
forecasts <- NA
```

```
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  dtotaldatappieds <- ts(dtotaldatappied[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaldatappieds, cor=TRUE)
  m<- POETKhat(dtotaldatappieds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  model <- auto.arima(PPIs,xreg=factors,ic=c("aic"))
  forecastm <- predict(model, n.ahead=6,newxreg=factors[i])
  forecastmt <- forecastm$pred[6]
  forecasts[i+1] <- forecastmt
}
FARIMAppi6 <- ts(forecasts, start=c(1997,07), frequency=12)
FARIMAppi6
# now doing FAVAR for PPI
T <- 205
start <- 103
forecasts <- NA
for (i in start:T){
  PPIs<- ts(dadPPI[1:i], start=c(1997, 02), frequency=12)
  dtotaldatappieds <- ts(dtotaldatappied[1:i,],
  start=c(1997,02),frequency=12)
  pca1 <- princomp(dtotaldatappieds, cor=TRUE)
  m<- POETKhat(dtotaldatappieds)
  factors <- ts(pca1$scores[1:i,1:m$K1BN],
  start=c(1997,02),frequency=12)
  varipdata <- cbind(PPIs,factors)
  model <- VAR(varipdata,p=1,type=c("const"),ic=c("AIC"))
```

```
  forecastm <- forecast(model,h=6)
  forecastmt <- forecastm$mean$PPIs[6]
  forecasts[i+1] <- forecastmt
}
FAVARppi6 <- ts(forecasts, start=c(1997,07), frequency=12)
FAVARppi6
# now writing user defined functions to compute MSE and AME
mse <- function(x,y,n){
  sum((x-y)^2)/n
}
n <- 98 # this is total number of out-of-samle period


cpiAR6 <- mse(dADcpi[109:206],cpiar6[109:206],n1)
cpiARIMA6 <- mse(dADcpi[109:206],cpiarima6[109:206],n)
cpiES6 <- mse(dADcpi[109:206],cpies6[109:206],n)
cpiDI6 <- mse(dADcpi[109:206],dicpi6[109:206],n)
cpiFARIMA6 <- mse(dADcpi[109:206],FARIMAcpi6[109:206],n)
cpiFAVAR6 <- mse(dADcpi[109:206],FAVARcpi6[109:206],n)
CPIforecastMSE6 <-cbind(cpiAR6,cpiARIMA6,cpiES6,cpiDI6,
cpiFARIMA6,cpiFAVAR6)
CPIforecastMSE6
# now for IP
ipAR6 <- mse(ldIP[109:206],arip6[109:206],n)
ipARIMA6 <- mse(ldIP[109:206],iparima6[109:206],n)
ipES6 <- mse(ldIP[109:206],ipes6[109:206],n)
ipDI6 <- mse(ldIP[109:206],diip6[109:206],n)
ipFARIMA6 <- mse(ldIP[109:206],FARIMAip6[109:206],n)
ipFAVAR6 <-mse(ldIP[109:206],FAVARIP6[109:206],n)
IPforecastMSE6 <-cbind(ipAR6,ipARIMA6,ipES6,ipDI6,ipFARIMA6,ipFAVAR6)
```

```
IPforecastMSE6
# now for Production of electricity
peAR6 <- mse(dladPE[109:206],pear6[109:206],n)
peARIMA6 <- mse(dladPE[109:206],pearima6[109:206],n)
peES6 <- mse(dladPE[109:206],pees6[109:206],n)
peDI6 <- mse(dladPE[109:206],dipe6[109:206],n)
peFARIMA6 <- mse(dladPE[109:206],FARIMApe6[109:206],n)
peFAVAR6 <- mse(dladPE[109:206],FAVARpe6[109:206],n)
PEforecastMSE6<-cbind(peAR6,peARIMA6,peES6,peDI6,peFARIMA6,peFAVAR6)
PEforecastMSE6
#now for PPI
ppiAR6 <- mse(dadPPI[109:206],ppiar6[109:206],n)
ppiARIMA6 <- mse(dadPPI[109:206],ppiarima6[109:206],n)
ppiES6 <- mse(dadPPI[109:206],ppies6[109:206],n)
ppiDI6 <- mse(dadPPI[109:206],dippi6[109:206],n)
ppiFARIMA6 <- mse(dadPPI[109:206],FARIMAppi6[109:206],n)
ppiFAVAR6 <- mse(dadPPI[109:206],FAVARppi6[109:206],n)
PPIforecastMSE6 <- cbind(ppiAR6,ppiARIMA6,ppiES6,
ppiDI6,ppiFARIMA6,ppiFAVAR6)
CPIforecastMSE6
IPforecastMSE6
PEforecastMSE6
PPIforecastMSE6
totalforecast6 <-rbind(CPIforecastMSE6,
IPforecastMSE6,PEforecastMSE6,PPIforecastMSE6)
totalforecast6
colnames(totalforecast6) <- c("AR","ARIMA","ES","DI",
"FARIMA","FAVAR")
rownames(totalforecast6) <- c("CPI","IP","PE","PPI")
```

```
totalMSE6 <-t(totalforecast6)
totalMSE6

#now doing a relatively MSE and MFE
RCPImse6 <- totalMSE6[,1]/totalMSE6[1,1]
RIPmse6 <- totalMSE6[,2]/totalMSE6[1,2]
RPEmse6 <- totalMSE6[,3]/totalMSE6[1,3]
RPPImse6 <-totalMSE6[,4]/totalMSE6[1,4]
RCPImse6
RIPmse6
RPEmse6
RPPImse6
rbind(RCPImse6,RIPmse6,RPEmse6,RPPImse6)
RtotalMSE6 <- t(rbind(RCPImse6,RIPmse6,RPEmse6,RPPImse6))
colnames(RtotalMSE6) <- c("CPI","IP","PE","PPI")
RtotalMSE6

totalMAFE6 <- xtable(RtotalMAE6)
totalMSFE6 <- xtable(RtotalMSE6)
print.xtable(totalMAFE6,type="latex", file="",floating=TRUE,table.placement="H")
print.xtable(totalMSFE6,type="latex", file="",floating=TRUE,table.placement="H")
tsCPI6 <- ts(dADcpi[109:206],start=c(2006,02),frequency=12)
tscpiar6 <-ts(cpiar6[109:206],start=c(2006,02),frequency=12)
tscpiarima6 <-ts(cpiarima6[109:206],start=c(2006,02),frequency=12)
tscpies6 <-ts(cpies6[109:206],start=c(2006,02),frequency=12)
tscpidi6 <-ts(dicpi6[109:206],start=c(2006,02),frequency=12)
tscpiFARIMA6 <-ts(FARIMAcpi6[109:206],start=c(2006,02),frequency=12)
tscpiFAVAR6 <-ts(FAVARcpi6[109:206],start=c(2006,02),frequency=12)
ts.plot(tsCPI6,tscpiar6,
```

```
        gpars=list(main="plot of CPI AR forecast and actual
        value",xlab="time",ylab="difference of CPI",lty=c(1:2)))
ts.plot(tsCPI6,tscpiarima6,
        gpars=list(main="plot of CPI ARIMA forecast and actual
        value",xlab="time",ylab="difference of CPI",lty=c(1:2)))
ts.plot(tsCPI6,tscpies6,
        gpars=list(main="plot of CPI EM forecast and actual
        value",xlab="time",ylab="difference of CPI",lty=c(1:2)))
ts.plot(tsCPI6,tscpidi6,
        gpars=list(main="plot of CPI DI forecast and actual
        value",xlab="time",ylab="difference of CPI",lty=c(1:2)))
ts.plot(tsCPI6,tscpiFARIMA6,
        gpars=list(main="plot of CPI FARIMA forecast and actual
        value",xlab="time",ylab="difference of CPI",lty=c(1:2)))
ts.plot(tsCPI6,tscpiFAVAR6,
        gpars=list(main="plot of CPI FAVAR forecast and actual
        value",xlab="time",ylab="difference of CPI",lty=c(1:2)))
tsIP6 <- ts(ldIP[109:206],start=c(2006,02),frequency=12)
tsipar6 <-ts(arip6[109:206],start=c(2006,02),frequency=12)
tsiparima6 <-ts(iparima6[109:206],start=c(2006,02),frequency=12)
tsipes6 <-ts(ipes6[109:206],start=c(2006,02),frequency=12)
tsipdi6 <-ts(diip6[109:206],start=c(2006,02),frequency=12)
tsipFARIMA6 <-ts(FARIMAip6[109:206],start=c(2006,02),frequency=12)
tsipFAVAR6 <-ts(FAVARIP6[109:206],start=c(2006,02),frequency=12)
ts.plot(tsIP6,tsipar6,
        gpars=list(main="plot of IP AR forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsIP6,tsiparima6,
        gpars=list(main="plot of IP ARIMA forecast and actual
```

```
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsIP6,tsipes6,
        gpars=list(main="plot of IP EM forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsIP6,tsipdi6,
        gpars=list(main="plot of IP DI forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsIP6,tsipFARIMA6,
        gpars=list(main="plot of IP FARIMA forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsIP6,tsipFAVAR6,
        gpars=list(main="plot of IP FAVAR forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))

tsPE6 <- ts(dladPE[109:206],start=c(2006,02),frequency=12)
tspear6 <-ts(pear6[109:206],start=c(2006,02),frequency=12)
tspearima6 <-ts(pearima6[109:206],start=c(2006,02),frequency=12)
tspees6 <-ts(pees6[109:206],start=c(2006,02),frequency=12)
tspedi6 <-ts(dipe6[109:206],start=c(2006,02),frequency=12)
tspeFARIMA6 <-ts(FARIMApe6[109:206],start=c(2006,02),frequency=12)
tspeFAVAR6 <-ts(FAVARpe6[109:206],start=c(2006,02),frequency=12)
ts.plot(tsPE6,tspear6,
        gpars=list(main="plot of PE AR forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPE6,tspearima6,
        gpars=list(main="plot of PE ARIMA forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPE6,tspees6,
        gpars=list(main="plot of PE EM forecast and actual
```

```
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPE6,tspedi6,
        gpars=list(main="plot of PE DI forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPE6,tspeFARIMA6,
        gpars=list(main="plot of PE FARIMA forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPE6,tspeFAVAR6,
        gpars=list(main="plot of PE FAVAR forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))

tsPPI6 <- ts(dadPPI[109:206],start=c(2006,02),frequency=12)
tsppiar6 <-ts(ppiar6[109:206],start=c(2006,02),frequency=12)
tsppiarima6 <-ts(ppiarima6[109:206],start=c(2006,02),frequency=12)
tsppies6 <-ts(ppies6[109:206],start=c(2006,02),frequency=12)
tsppidi6 <-ts(dippi6[109:206],start=c(2006,02),frequency=12)
tsppiFARIMA6 <-ts(FARIMAppi6[109:206],start=c(2006,02),frequency=12)
tsppiFAVAR6 <-ts(FAVARppi6[109:206],start=c(2006,02),frequency=12)
ts.plot(tsPPI6,tsppiar6,
        gpars=list(main="plot of PPI AR forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPPI6,tsppiarima6,
        gpars=list(main="plot of PPI ARIMA forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPPI6,tsppies6,
        gpars=list(main="plot of PPI EM forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPPI6,tsppidi6,
        gpars=list(main="plot of PPI DI forecast and actual
```

```
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPPI6,tsppiFARIMA6,
        gpars=list(main="plot of PPI FARIMA forecast and actual
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
ts.plot(tsPPI6,tsppiFAVAR6,
        gpars=list(main="plot of PPI FAVAR forecast and actual \
        value",xlab="time",ylab="Percentage change",lty=c(1:2)))
```

Bibliography

(2012). China 2030: Building a modern, harmonious, and creative high-income society.

(2014). *Australian National Accounts: National Income, Expenditure and Product, Jun 2014*. Australian Bureau of Statistics. http://www.abs.gov.au/ausstats/abs@.nsf/mf/5206.0/.

(2014). *Trade and economic statistics*. Australian Government: Department of Foreign Affairs and Trade. http://www.dfat.gov.au/trade/trade-and-economic-statistics.html.

Adler, S. (2013). *The Chinese Economy*, volume 1. Routledge.

Angelini, E., Henry, J., and Mestre, R. (2001). Diffusion index-based inflation forecasts for the euro area1. In *Papers presented*, page 109.

Artis, M. J., Banerjee, A., and Marcellino, M. (2005). Factor forecasts for the uk. *Journal of Forecasting*, 24(4):279–298.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2008). *Large dimensional factor analysis*. Now Publishers Inc.

Banerjee, A., Marcellino, M., and Masten, I. (2005). Leading indicators for euro-area inflation and gdp growth*. *Oxford Bulletin of Economics and Statistics*, 67(s1):785–813.

Basilevsky, A. T. (2009). *Statistical factor analysis and related methods: theory and applications*, volume 418. John Wiley & Sons.

Bennett, J. and Dixon, H. D. (1996). A macrotheoretic model of the chinese economy. *Journal of Comparative Economics*, 22(3):277–294.

Bernanke, B. S. and Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3):525–546.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2004). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. Technical report, National Bureau of Economic Research.

Boivin, J. and Ng, S. (2005). Understanding and comparing factor-based forecasts. Technical report, National Bureau of Economic Research.

Breitung, J. and Eickmeier, S. (2006). Dynamic factor models. *Allgemeines Statistisches Archiv*, 90(1):27–42.

Brisson, M., Campbell, B., and Galbraith, J. W. (2003). Forecasting some low-predictability time series using diffusion indices. *Journal of Forecasting*, 22(6-7):515–531.

Camacho, M. and Sancho, I. (2003). Spanish diffusion indexes. *Spanish economic review*, 5(3):173–203.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research*, 295(2):295–336.

Chow, G. (2006). Are chinese official statistics reliable? *CESifo Economic Studies*, 52(2):396–414.

Chow, G. C. (1985). A model of chinese national income determination. *The Journal of Political Economy*, pages 782–792.

Chow, G. C. (1986). Chinese statistics. *The American Statistician*, 40(3):191–196.

Chow, G. C. (2010). Note on a model of chinese national income determination. *Economics Letters*, 106(3):195–196.

Cristadoro, R., Forni, M., Reichlin, L., and Veronese, G. (2001). *A core inflation index for the euro area.* Centre for Economic Policy Research.

Cristadoro, R., Forni, M., Reichlin, L., and Veronese, G. (2005). A core inflation indicator for the euro area. *Journal of Money, Credit and Banking*, pages 539–560.

den Reijer, A. H. (2005). Forecasting dutch gdp using large scale factor models. Technical report, Netherlands Central Bank, Research Department.

DAgostino, A., Giannone, D., and Surico, P. (2006). (un) predictability and macroeconomic stability. Technical report.

Fernald, J., Spiegel, M. M., and Swanson, E. T. (2014). Monetary policy effectiveness in china: evidence from a favar model. *Journal of International Money and Finance.*

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, 82(4):540–554.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2003). Do financial variables help forecasting inflation and real activity in the euro area? *Journal of Monetary Economics*, 50(6):1243–1255.

Giannone, D. and Matheson, T. (2007). A new core inflation indicator for new zealand.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Gosselin, M.-A. and Tkacz, G. (2001). *Evaluating factor models: an application to forecasting inflation in Canada*. Bank of Canada.

Graff, M., Matheson, T., et al. (2004). Composite leading indicators of inflation in new zealand. Technical report, Citeseer.

Gupta, R. and Kabundi, A. (2010). Forecasting macroeconomic variables in a small open economy: a comparison between small-and large-scale models. *Journal of Forecasting*, 29(1-2):168–185.

Gupta, R. and Kabundi, A. (2011). A large factor model for forecasting macroeconomic variables in south africa. *International Journal of Forecasting*, 27(4):1076–1088.

Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617.

Holz, C. A. (2003). fast, clear and accurate: How reliable are chinese output and economic growth statistics? *The China Quarterly*, 173:122–163.

Holz, C. A. (2004). China's statistical system in transition: Challenges, data problems, and institutional innovations. *Review of Income and Wealth*, 50(3):381–409.

Holz, C. A. (2008). China's 2004 economic census and 2006 benchmark revision of gdp statistics: more questions than answers? *The China Quarterly*, 193:150–163.

Holz, C. A. and Lin, Y.-m. (2001a). The 1997–1998 break in industrial statistics: Facts and appraisal. *China Economic Review*, 12(4):303–316.

Holz, C. A. and Lin, Y.-m. (2001b). Pitfalls of china's industrial statistics: inconsistencies and specification problems. *China Review*, pages 29–71.

John, G. (1977). The dynamic factor analysis of economic time series.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327.

Lin, C.-Y. and Wang, C. (2013). Forecasting china's inflation in a data-rich environment. *Applied Economics*, 45(21):3049–3057.

Liu, B. *Development and Application of Econometric Models by the Central Banks*.

Liu, D. and Jansen, D. W. (2007). Macroeconomic forecasting using structural factor analysis. *International Journal of Forecasting*, 23(4):655–677.

Lubik, T. A. and Schorfheide, F. (2007). Do central banks respond to exchange rate movements? a structural investigation. *Journal of Monetary Economics*, 54(4):1069–1087.

Marcellino, M., Stock, J. H., and Watson, M. W. (2003). Macroeconomic forecasting in the euro area: Country specific versus area-wide information. *European Economic Review*, 47(1):1–18.

Mehrotra, A. and Sánchez-Fung, J. R. (2008). Forecasting inflation in china. *China Economic Journal*, 1(3):317–322.

Moser, G., Rumler, F., and Scharler, J. (2007). Forecasting austrian inflation. *Economic Modelling*, 24(3):470–480.

Nakamura, E., Steinsson, J., and Liu, M. (2014). Are chinese growth and inflation too smooth? evidence from engel curves. Technical report, National Bureau of Economic Research.

Osborne, J. W. and Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical assessment, research & evaluation*, 9(6):1–12.

Qin, D., Cagas, M. A., Ducanes, G., He, X., Liu, R., Liu, S., Magtibay-Ramos, N., and Quising, P. (2007). A macroeconometric model of the chinese economy. *Economic Modelling*, 24(5):814–822.

Rawski, T. G. (1976). On the reliability of chinese economic data.

Rawski, T. G. and Mead, R. W. (1998). On the trail of china's phantom farmers. *World Development*, 26(5):767–781.

Reichlin, L. (2002). Factor models in large cross-sections of time series.

Samuelson, P. A. and Nordhaus, W. D. (1998). Economics. boston, ma, burr ridge, il, dubuque, ia, etc.

Sargent, T. J., Sims, C. A., et al. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research*, 1:145–168.

Schumacher, C. (2007). Forecasting german gdp using alternative factor models based on large datasets. *Journal of Forecasting*, 26(4):271–302.

Schumacher, C. and Breitung, J. (2008). Real-time forecasting of german gdp based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, 24(3):386–398.

Schumacher, C. and Dreger, C. (2002). Estimating large-scale factor models for economic activity in germany: Do they outperform simpler models?

Stock, J. H. and Watson, M. W. (1998). Diffusion indexes. Technical report, National bureau of economic research.

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44(2):293–335.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stock, J. H. and Watson, M. W. (2006). Forecasting with many predictors. *Handbook of economic forecasting*, 1:515–554.

Stock, J. H. and Watson, M. W. (2011). Dynamic factor models. *Oxford Handbook of Economic Forecasting. Oxford University Press, USA*, pages 35–59.

Van Nieuwenhuyze, C. (2005). A generalized dynamic factor model for the belgian economy: identification of the business cycle and gdp growth forecasts. *Journal of Business Cycle Measurement and Analysis*, 2005(2):213–247.

Wold, H. (1985). Partial least squares. *Encyclopedia of statistical sciences*.

Zheng, T. and Guo, H. (2013). Estimating a small open economy dsge model with indeterminacy: Evidence from china. *Economic Modelling*, 31:642–652.

Zhou, R., Wang, X., and Tong, G. (2013). Forecasting macroeconomy based on the term structure of credit spreads: evidence from china. *Applied Economics Letters*, 20(15):1363–1367.