BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2011-11

Mixed Frequency Forecasts for Chinese GDP

by Philipp Maier

Bank of Canada Working Paper 2011-11

April 2011

Mixed Frequency Forecasts for Chinese GDP

by

Philipp Maier

International Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
pmaier@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

I would like to thank Benoît-Nicolas Houdet-Côte for information on Chinese data and Sharon Kozicki, Marco Lombardi, Tatevik Sekhposyan and Jamie Hall, as well as seminar participants at the Bank of Canada and the Reserve Bank of Australia, for helpful comments. Golbahar Kazemian, Stephen Hamza and William Wu provided excellent research assistance.

Abstract

We evaluate different approaches for using monthly indicators to predict Chinese GDP for the current and the next quarter (‘nowcasts’ and ‘forecasts’, respectively). We use three types of mixed-frequency models, one based on an economic activity indicator (Liu et al., 2007), one based on averaging over indicator models (Stock and Watson, 2004), and a static factor model (Stock and Watson, 2002). Evaluating all models’ out-of-sample projections, we find that all the approaches can yield considerable improvements over naïve AR benchmarks. We also analyze pooling across forecasting methodologies. We find that the most accurate nowcast is given by a combination of a factor model and an indicator model. The most accurate forecast is given by a factor model. Overall, we conclude that these models, or combinations of these models, can yield improvements in terms of RMSE’s of up to 60 per cent over simple AR benchmarks.

JEL classification: C50, C53, E37, E47
Bank classification: Econometric and statistical methods; International topics

Résumé

L’auteur évalue différentes approches fondées sur l’emploi d’indicateurs mensuels pour prévoir le PIB chinois pour le trimestre courant et le trimestre à venir. Il a recours à trois techniques d’estimation à fréquence mixte : la première est basée sur un indicateur de l’activité économique (Liu et autres, 2007); la deuxième utilise la moyenne des valeurs calculées au moyen de différents modèles indicateurs (Stock et Watson, 2004); la dernière fait appel à un modèle factoriel statique (Stock et Watson, 2002). D’après les résultats qu’il obtient, chacune de ces approches peut produire des projections hors échantillon bien meilleures que des modèles autorégressifs simples. L’auteur examine également si le fait de combiner les méthodes de prévision offre des avantages. Il constate que la prévision la plus exacte pour le trimestre courant résulte de la combinaison d’un modèle factoriel et d’un modèle indicateur. La meilleure prévision pour le trimestre à venir est tirée d’un modèle factoriel. L’auteur conclut globalement que ces modèles, ou des combinaisons de ceux-ci, donnent lieu à des réductions pouvant aller jusqu’à 60 % de la racine de l’erreur quadratique moyenne par rapport à des modèles autorégressifs simples.

Classification JEL : C50, C53, E37, E47
Classification de la Banque : Méthodes économétriques et statistiques; Questions internationales

1 Introduction

Forecasting models for China are hard to come by. In general terms, forecasting entails selecting a set of predictors, and choosing a functional form and estimation method to map this information into the forecasts (Timmermann, 2005). When forecasting China's economic growth, however, several additional challenges arise: The relatively poor quality of Chinese data can compromise the selection of predictors, the short sample period, over which many indicators are available, complicates estimation, and the possibility that the rapid transformation of the economy leads to structural breaks makes it difficult to select a suitable functional form. In light of these complications, it is likely that the best way to exploit the information content of Chinese statistics is not by attempting to identify one single dominant forecasting method, but rather to use a combination of data or forecasts. Over the past years, forecasting using many predictors has become popular, with methods including factor models and forecast pooling (Groen and Kapetanios, 2008). However, most of these methods have not yet been applied to the Chinese economy.[1] Given China's rapidly growing importance for the global economy – including commodity prices – better tools to predict the state of the Chinese economy are important to help improve our understanding of the global business cycle.

To fill this gap in the literature, this study proposes and evaluates short-term forecasting models for China. Specifically, we use monthly indicators to project GDP for the current and for the next quarter ('nowcast' and 'forecast', respectively), using three different methodologies. First, an economic activity indicator in the spirit of the Conference Board (2001), based on a study by Liu et al. (2007). This indicator combines monthly series of economic activity, mixing supply and demand side elements. The second approach explores the merits of forecast averaging, using 33 monthly indicators. Following Stock and Watson's (2004) insight that pooling projections can substantially improve forecasting accuracy, we estimate several versions of indicator models (with lagged GDP or lagged indicators), optimally select the best specification, and project each equation. These individual out-of-sample forecasts are subsequently aggregated, yielding a 'composite forecast'. Third, we estimate a static factor model following Stock and Watson (2002). Moreover, once we have estimated and evaluated all three approaches individually, we also examine various ways to mix these methodologies and pool forecasts across different forecasting methods.

Evaluating out-of-sample-forecasts for the period 2008Q2-2010Q4, we find that all approaches can considerably outperform naive AR benchmarks. Improvement in terms of accuracy can be substantial, in particular for the factor model, which reduces root mean squared errors by up to 50 per cent for the nowcast and almost 60 per cent for the forecast, relative to the AR benchmark.

[1] In fact, very few papers attempt to model or forecast the Chinese economy. An important exception is the 'Global Projection Model' (GPM), which was originally developed by the IMF (Carabenciov et al., 2008) and adapted for the Bank of Canada in Bailliu at al. (2010). Note, however, that GPM is a long-term forecasting model, whereas we examine short-term forecasts.

We also find that pooling forecasts across methodologies is a useful way to improve accuracy even further, notably early in the quarter. Overall, we interpret these results as underlining the potential of forecast pooling for the Chinese economy. The methods we propose provide some safeguard against structural breaks or errors in the data, since they use a relatively large data set and individual forecasts are thus less dependent on the evolution of a single indicator. On this basis, we conclude that forecast averaging, in particular when including a factor model, seems a promising technique for short-term forecasting of the Chinese economy.

We proceed as follows. In section 2, we briefly discuss issues surrounding the quality of Chinese data, and review our selection of indicators. In section 3, we outline each forecasting approach individually, before presenting the results in section 4. In section 5, we pool forecasts across methodologies. The final section discusses the relative merits of all approaches, and offers some ideas for future research.

2 Data sources

In this study we forecast China's official GDP releases for the current and next quarter.[2] In this context, several complications arise.

First, the reliability of Chinese data has been debated fairly extensively. In a well-known study, Maddison (1998) challenges official Chinese GDP data, and suggests that official statistics overstated GDP growth by almost 2.4 percentage points for the years 1952 through 1995. In a similar vein, Rawski (2001b) argues that official Chinese statistics contain major exaggerations of real output growth and that standard data contains numerous inconsistencies.[3] Zheng (2001) notes 'serious weakness' in some fields of national accounts, with quarterly GDP estimates being 'crudely calculated with heavy reliance on estimates and excessive aggregation'. In contrast, Holz (2005) and Chow (1986) note that while not all individual data points might be accurate, systematic data falsification at the higher levels of the statistical bureaucracy is unlikely,[4] and longer-term trends are generally found to be fairly accurate. Using factor analysis to construct a coincident indicator, Mehrotra and Pääkkönen (2011) conclude that for the most part of their sample (1997-2009), GDP matches the dynamics of the coincident indicator. This suggests that official data seems to provide a reasonably reliable

[2] China's official GDP is typically released very fast; for instance, 2010Q2 was released 15 days after quarter end. Given the short publication lags, we do not backcast Chinese GDP.

[3] The literature on Chinese statistics uses two procedures to verify official data publications. First, it compares output growth to other variables which one would expect to be correlated with output (like the growth rate of energy use or freight data, see Lian and Xiaolu, 2000). A common finding is that since 1992, in particular, industrial value-added grew more than other indicators of economic growth. Second, there are multiple ways to construct aggregate nation-wide GDP data. Official statistics rely on the production approach, while aggregating the sum of incomes or the sum of expenditures suggest a lower level of GDP growth (see Rawski, 2001a).

[4] However, Holz (2005) notes that falsification of data in the countryside is beyond doubt.

picture, discrepancies at specific points of time notwithstanding (for instance during the Asian crisis).[5]

Since our objective is to forecast official Chinese GDP releases, we are less concerned about whether or not the level of GDP is an accurate representation of Chinese economy activity (we simply take as given that Chinese data might be overstating 'true' growth). However, even within this narrowly defined context, some data issues remain, as the following examples show.

The first issue is that China reports three different measures for GDP: quarterly annualized GDP growth rates, GDP growth since the beginning of the year at quarterly frequency (also known as 'year-to-date'), and annual GDP growth rates.[6] Unfortunately, they are not always identical, and the quarterly numbers may or may not add up to the annual growth rate. For instance, the official annual GDP growth rate for 2009 is 9.1 per cent, yet calculating an annual GDP growth rate using the quarterly GDP releases would yield 8.4 per cent.[7] Also, it is not uncommon that the annual growth figure is revised, while the official quarterly numbers for the same year remain unchanged. In this study, we compare quarterly GDP forecasts, but if the models were extended to yield annual GDP projections, a decision would need to be taken whether to forecast the official annual number, or rather each quarter individually, while acknowledging that quarterly growth rates need not add up to the annual figure.

Second, some of the indicators used to forecast GDP contain discrepancies when series are published in both growth rates and levels. Figure 1 shows retail sales, both in terms of the official growth rate of retail sales, and the growth rates calculated off the series in levels. We would expect the two to be identical (up to small rounding errors), but the differences are actually substantial; for instance, in February, April and May of 2005 and in February 2010, the difference exceeded 10 percentage points.[8] This complicates forecasting, because on the one hand, the level series can more easily be deflated to compute an index of real retail sales; on the other hand, most forecasters will likely focus on the National Bureau of Statistics' release of yearly growth of retail sales, so consistency with external analysis might require using the yearly growth rates.

Lastly, some indicators are not very reliable, reducing their value for forecasting purposes. Chinese house price data is particularly challenging. Using monthly property price growth rates[9] to create a index of property prices, the series suggests that over the past five years, property prices in Shanghai rose only by 10 per cent in real terms, and that Beijing prices are only 7 per cent higher than the national average. Numerous anecdotal evidence on the Chinese real estate market suggests that Chinese property prices have risen substantially

[5]The overall score for China's data in the World Bank Statistical Capacity Indicator was 58 on a scale of 0-100 (compared to an average of 65 of all countries included).

[6]Officially, quarter-on-quarter Chinese GDP growth is not available, but Abeysinghe and Rajaguru (2004) outline a methodology to estimate them.

[7]Since 2000Q1 – the first year for which this comparison is possible – the quarterly series underestimates the annual GDP report, on average, by 0.7 percentage points, but discrepancies can be as large as 1.5 percentage points (recorded in 2006).

[8]The differences are likely due to revisions to the level, while growth rates were not revised.

[9]The series started in 2005 and covers 70 Chinese cities.

Figure 1: Growth rates of retail sales (y/y) as published by the NBS, and calculated off the level of retail sales

faster, and that Beijing is considerably more expensive.[10]

Taken together, we believe that longer-term trends in indicators of economic activity are generally found to provide a fairly accurate picture. However, during periods of high volatility or in areas of particular attention – like Chinese property markets – the bias in official statistics could be large. This poses an important additional challenge for short-term forecasting, if estimates of economic activity put a high weight on latest data points.

We deal with these issues as follows. To provide some safeguard against structural breaks, we combine forecasts from many different models. To (partially) alleviate concerns about data quality and errors in single economic indicators, our approach is to forecast in a data-rich environment, i.e. to use many different indicators. Our starting point is the data required for the China economic activity indicator, but for the factor model and indicator model averaging, we expand the data set with other indicators. We select domestic indicators based on data availability, as well as timeliness of publication. Given the importance of the export sector for the Chinese economy, we also add a number of indicators to proxy for foreign demand, as well as the state of the global economy (we include the PMI Manufacturing for the euro area, Japan, the United States, and the Global PMI manufacturing).

Table 1 contains a description of the series we use. They cover all important sectors of the economy, including monetary aggregates and bank lending data; production-, demand- and trade-related data, consumer confidence, and stock market indices. All series are obtained from CEIC, unless indicated otherwise. Data released as year-to-date has been transformed into yearly growth rates, and all series published in levels have been transformed into yearly growth rates to ensure stationarity.[11]

[10] One way to gain a likely more accurate picture is to divide the total value of properties sold divided by total squared meter sold. The resulting index seems to be more in line with anecdotal evidence.

[11] In this study, we did not explicitly consider publication lags, i.e. suppressing information

Table 1: List of Chinese variables

Variable (sources in brackets)	Included in the following models	Publication lag
GDP	China economic activity index, Indicator model averaging, Factor model	21 days
Fixed Asset Investment (Global Insight)	China economic activity index, Indicator model averaging	15 days
Fixed Asset Investment (CIEC)	Indicator model averaging	15 days
Industrial Production	China economic activity index, Indicator model averaging	11 days
Electricity Production	China economic activity index, Indicator model averaging	19 days
Energy Production	Indicator model averaging	18 days
Exports	China economic activity index, Indicator model averaging, Factor model	10 days
Imports	Indicator model averaging, Factor model	10 days
Retail Sales of Consumer Goods	China economic activity index, Indicator model averaging	11 days
Consumer Confidence Index	China economic activity index (*) , Indicator model averaging, Factor model	24 days
Consumer Price Index	Indicator model averaging	11 days
Producer Price Index	Indicator model averaging, Factor model	13 days
Producer Price Index for Consumer Goods	Indicator model averaging, Factor model	13 days
Money Supply M0	Indicator model averaging, Factor model	19 days
Money Supply M2	Indicator model averaging, Factor model	19 days
PMI Manufacturing (NBS)	Indicator model averaging	1 day
PMI Manufacturing (Markit)	Indicator model averaging	5 days
PMI Manufacturing – New Orders (NBS)	Indicator model averaging	1 day
PMI Manufacturing – New Orders (Markit)	China economic activity index (*) , Indicator model averaging	5 days
Shanghai Stock Exchange Composite Index	Indicator model averaging, Factor model	0 days
Shenzhen Stock Exchange Composite Index	Indicator model averaging, Factor model	0 days
Industrial Sales	Indicator model averaging	13 days
Industrial Sales – Delivery Value for Export	Indicator model averaging	13 days
New Bank Loans	Indicator model averaging	11 days
Passenger traffic (km carried)	China economic activity index, Indicator model averaging, Factor model	20 days
Freight traffic (tons per km carried)	China economic activity index, Indicator model averaging, Factor model	20 days
Floor Space Started	Indicator model averaging, Factor model	24 days
Steel Production	Indicator model averaging, Factor model	13 days
Auto Production	Indicator model averaging, Factor model	17 days
Employment (total number of employees)	Indicator model averaging	50 days
Wage growth (total average remuneration, YTD)	Indicator model averaging	50 days
PMI Manufacturing EMU (Markit)	Indicator model averaging, Factor model	1 day
PMI Manufacturing Japan (Markit)	Indicator model averaging, Factor model	1 day
PMI Manufacturing United States (ISM)	Indicator model averaging, Factor model	1 day
PMI Manufacturing Global (Markit)	Indicator model averaging, Factor model	1 day

(*) Not included in Liu et al. (2007), see section 3.2.

3 Methodology

We evaluate projections of three different types of models, as well as various combinations of these models. As a benchmark, we chose a simple AR benchmark model.[12] We focus on short-term forecasting, or, more specifically, on 'nowcasting', defined as a projection for the current quarter, and a one-quarter ahead 'forecast'.[13] While the benchmark model is estimated using quarterly data, the other models employ mixed frequency techniques to obtain monthly GDP estimates. To project monthly indicators onto quarterly data, we use a bridge equation and assume the following.[14]

- In the first month, the quarterly value is identical to the value recorded in the first month;
- In the second month, the value in the missing month of the quarter will be identical to the last value.

3.1 The AR benchmark model

Formally, the simple AR benchmark model is given by:

$$y_{t+h} = \alpha + \sum_k \beta_k^h y_{t-k} + e_{t+h}, \tag{1}$$

where for each out-of-sample forecast, the optimal lag length k is determined by the Schwarz criterion. Unlike the other models, the AR benchmark is estimated using quarterly data. Figure 2 shows actual Chinese real GDP and out-of-sample nowcasts and forecasts for the last three years, using the AR benchmark model.

3.2 China Economic activity indicator

Economic activity indicators have been developed for various economies, including the United States (Aruoba et al., 2009)[15] and the euro area (Camacho and Perez-Quiros, 2010). The methodologies behind these indicators of economic activity differ somewhat, but the general idea is to blend high- and low-frequency

---

for months where certain indicators would not have been available yet. However, as shown, only two series – employment and wage growth – are not published in the same month, so switching to pseudo-real time data will not affect our results substantially.

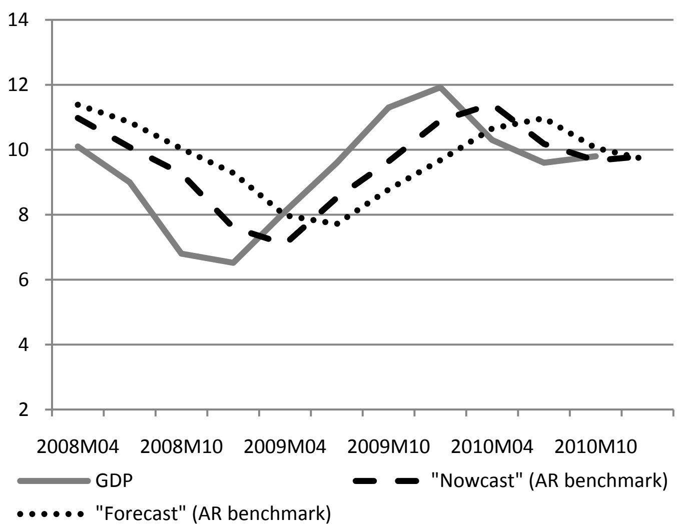
[12]AR models are standard benchmarks in the literature; note that a different benchmark would not change the ranking of our forecasting methods qualitatively.

[13]The term 'projection' covers both nowcasting and forecasting. Strictly speaking, the 'AR benchmark Nowcast' is a forecast, too, since no contemporaneous information is used (effectively a one- and two-period forecast). We still refer to the specification as 'nowcast' to simplify the comparison with the other two methods.

[14]This type of bridge equation has been found to deliver good forecasts (Diron, 2008).

[15]The Aruoba-Diebold-Scotti 'Business Conditions Index' is published weekly by the Federal Reserve Bank of Philadelphia.

Figure 2: Chinese real GDP growth and a simple AR benchmark model



Note: For consistency with the other models we refer to the 'Nowcast' for the current quarter forecast and the 'Forecast' for the next quarter forecast

information from different sources to obtain a continuous impression of the overall level economic activity. While Aruoba et al. (2009) are careful to point out that economic activity may only be loosely related to GDP – which implies that the Arouba-Diebold-Scotti 'Business Conditions Index' might not be the best way to nowcast or forecast U.S. GDP – Camacho and Perez-Quiros (2010) state explicitly that they view their (relatively similar) model as an important short-term forecasting tool.

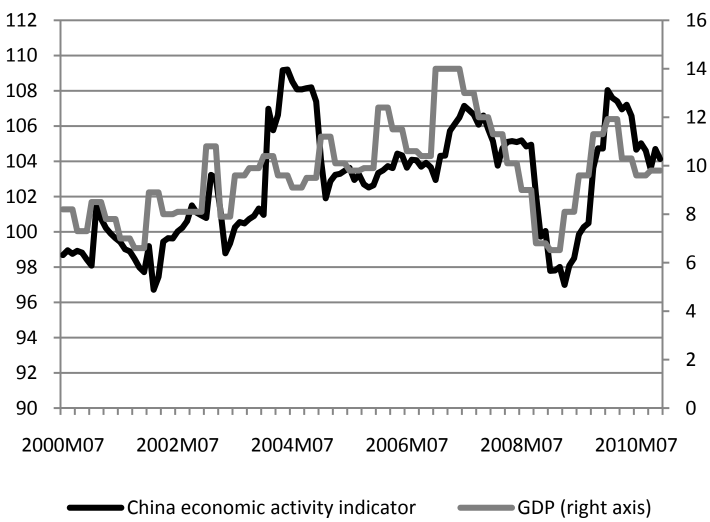
A related, albeit methodologically much simpler approach, is composite leading indicators, as developed by the Conference Board (2001) for the United States and the OECD (2008) for major industrialized economies.[16] Their purpose is to track series that – in the past – have shown to exhibit properties that can give early signals about the business cycle, in particular, signaling turning points. The general idea is to select series that precede the business cycle, filter them to remove seasonal factors, and aggregate them. The aggregation is often done using simple, unweighted averages, but more complex weighting schemes are possible, too.

In this spirit, the Hong Kong Monetary Authority (HKMA) developed an economic activity indicator for China (Liu et al., 2007).[17] Liu et al. (2007)

[16] In May 2010, the Conference Board also introduced a 'Leading Economic Index' for China, aimed at signaling turning points in China's economic cycles.

[17] Goldman Sachs (2004) uses a similar approach to construct a real activity index for China, but employs a different weighting scheme, based on the correlation between the indicators and

Figure 3: China real activity index and Chinese real GDP growth



rely on eight indicators of economic activity, combining supply side elements (industrial and electricity production) and demand side indicators (exports, real retail sales, real fixed asset investment, and volumes of passenger and freight transport) with real per-capita household income in urban areas. Then, the real activity index i is computed as a weighted average of month-to-month changes for each indicator x, with weights proportional to each indicator's standard deviation (the more volatile an indicator, the smaller its weight). Formally,

$$i_t = \sum_i w_i x_{it}, \text{ with } w_i = \frac{s_i}{\sum_i s_i}, \tag{2}$$

with s_i denoting the inverse of the standard deviation of indicator x.

When constructing this indicator, a complication is that China has switched from monthly to quarterly frequency for real per-capita household income in urban areas. Hence, this series is no longer available as a timely indicator. We include consumer confidence and the 'New orders' subcomponent of the Markit PMI Manufacturing index instead.

Figure 3 plots the China real activity index and real Chinese GDP. Clearly, the series are strongly correlated, and overall, the China real activity index seems to provide an accurate indication of future Chinese GDP growth. Interestingly, the correlation seems to be particularly tight during the slowing in economic activity in 2008/2009 and the subsequent recovery, while less accurate in the mid-2000s.

real GDP.

To forecast with the HKMA economic activity indicator, we estimate the following model:

$$y_{t+h} = \alpha + \sum_k \beta_k^h y_{t-k} + \sum_l \gamma_l^h HKMA_{t-l} + e_{t+h}, \tag{3}$$

and select the optimal number of lags for GDP (k) and the economic activity indicator (l) based on the Schwarz criterion.[18] In all cases, the preferred model specification contains 1 lag of the China economic activity indicator, and 1 or 2 lags of GDP.[19]

3.3 Forecast averaging using indicator models

While economic activity indices rely on a relatively small set of economic variables, a conceptually different approach is to employ a large data set, but to process it efficiently. Our next approach is to estimate a large number of indicator models, and pool forecasts to improve forecasting accuracy.[20] The theory of optimal linear forecast pooling goes back to Bates and Granger (1969), and the idea is that by pooling forecasts over different data, the pooled forecast uses more information and should thus be more efficient than any individual forecast. Several pooling methods can be used, from simple averages to weighted averages (with time-invariant or time-varying weights, based e.g. on forecasts errors of the previous period). The reasons for the empirical success of forecast combinations – notably for simple forecast averages – are not yet very well understood (Groen and Kapetanios, 2008; Yang, 2004), but as Stock and Watson (2004) show, forecast averaging yields surprisingly good results. Simple averaging over forecasts can in many cases outperform more complex weighting schemes (Stock and Watson, 2003), in many cases rivaling accuracy of more complex models (see also Marcellino, 2004).[21]

Following Stock and Watson (2004), we forecast with the following class of indicator models:

$$y_{t+h} = \alpha + \sum_k \beta_k^h y_{t-k} + \sum_l \gamma_l^h i_{t-l} + e_{t+h}, \tag{4}$$

where y_t denotes GDP and i_t denotes one of the indicators from table 1. For each indicator, we estimate two versions of equation (4), which we denote as model 1 and model 2.

[18] As with all mixed frequency models, we assume for the quarterly value that in the first month, the quarter is identical to the value in the first month, and that in the second month, the missing third month value is identical to the second month's.

[19] More specifically, 33 out of 36 nowcasts, the preferred specification contains one lag of GDP (2 lags of GDP in 3 cases); the forecast always contains 1 lag of GDP.

[20] Forecast pooling is discussed e.g. in Diebold and Lopez (1996), Newbold and Harvey (2002), and Hendry and Clements (2002).

[21] A downside of forecasting averaging is that forecasting accuracy might fall, if a very good model is combined with a set of poorly performing models. In that case, an appropriate weighting scheme would give very little weight, if any, to the poorly performing models.

- For model 1, we include the contemporaneous value of one indicator in each of the models, plus vary the number of lags k for GDP;
- For model 2, we vary the number of lags l for the indicators, but do not include any lagged values of GDP.[22]

We reestimate all models for every period and, as recommended by Diebold (2007), we use the Schwarz criterion to select the optimal number of lags for k or l (with a maximum of 5 lags). Once the optimal specification for each indicator is found, we use each model to nowcast and forecast GDP. Then, we take the unweighted or weighted averages to obtain GDP forecasts.

3.4 Factor model

Our last forecasting methodology is a factor model. Factor models are based on the idea that large datasets can be represented by a small number of components, which are sufficient to characterize the main features of the data.[23] Since Sargent and Sims (1977), factor models have been used for macroeconomic applications,[24] and their properties are well suited to address some of the concerns when working with Chinese data: first, they allow to process large volumes of data very efficiently. Second, by extracting information from many series, factor models have been found to compensate for deficiencies in single economic indicators, including measurement errors or structural breaks.

An important drawback of factor models is that one cannot give an economic interpretation to the 'factors'. This can complicate the economic interpretation of the forecast, as it is not always evident which specific piece of information induces a change in the factors, and thus in the forecast.

Formally, a factor model expresses a N-dimensional time series X_t as

$$X_t = \Lambda F_t + e_t, \tag{5}$$

where F_t is a K-dimensional multiple time series of factors (with $K \ll N$), Λ is a matrix of loadings, relating the factors to the observed time series, and e_t are idiosyncratic disturbances. Equation (5) is not a standard regression model, as the factors are unobservable variables and F_t has to be estimated. This can be accomplished consistently by using the first K principal components of the data, i.e. the first K eigenvectors of the variance-covariance matrix of X_t.

Stock and Watson (2002) complement equation (5) with an equation describing the evolution of the 'target' variable y_t:

[22] In theory, one could use any combination of lagged dependent variable and indicator; given the short sample period, we decided to vary one, while keeping the other constant. Additional tests reveal that combinations of lagged indicators and lagged GDP typically do not yield superior forecasting power.

[23] Factor models can be viewed as a parsimonious alternative to large VAR models. Modeling interrelations among a large set of variables in a VAR system is not feasible because of the so-called 'curse of dimensionality', i.e. the fact the number of parameters to estimate grows rapidly. Factor models overcome this limitation by reducing the dimensionality of the data.

[24] The use of factor models originated in the finance literature, where researchers are faced with (for instance) large cross-sections of stock returns.

$$y_{t+1} = \beta' F_t + \gamma(L) y_t + \epsilon_{t+1}, \tag{6}$$

where $\gamma(L)$ is a polynomial in the lag operator. Forecasts are given by equation (6), and h-step-ahead forecasts can be constructed as follows:

$$y_{t+h} = \beta_h' F_t + \gamma_h(L) y_t + \epsilon_{t+h}. \tag{7}$$

The model generated by equations (5) and (6) is commonly referred to as 'static factor model', as no parametric dynamics are imposed on the factors.[25] Static factor models have e.g. been used by Schumacher and Breitung (2006) to forecast German GDP, by Gosselin and Tkacz (2001) to forecast Canadian inflation, and by Perevalov and Maier (2010) to forecast U.S. GDP. For China, Klein and Mak (2005) and Mak (2009) use a factor model as one of their inputs to forecast several economic indicators, including GDP and inflation, over the next 2 quarters. Mehrotra and Pääkkönen (2011) estimate a static factor model, using 83 economic indicators, to investigate the quality of statistics in China.

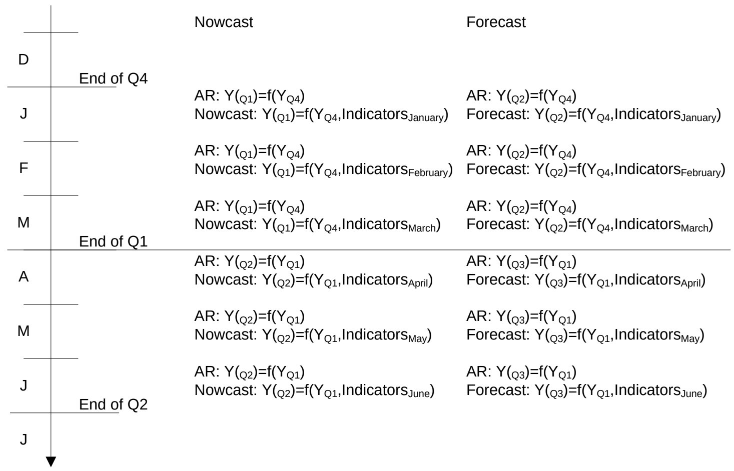
3.5 Forecast horizon

Figure 4 shows the timing of the forecasting exercise and the available data at each point in time. Suppose that we are in January. At this point, the AR benchmark model – our only quarterly model – 'nowcasts' Q1 GDP and forecasts Q2 GDP, based on Q4 GDP (plus possibly additional lags). All other models nowcast on the basis of Q4 GDP, plus information released in January. In February, all model forecasts are updated with February data (except for the AR model) etc.

We evaluate forecasting accuracy at the end of each month for the nowcast and next quarter's forecast. Our sample starts in 1999Q3 and ends in 2010Q4 and uses monthly data. All nowcasts and forecasts are done as rolling out-of-sample forecasts. Given the short sample horizon, we select as the forecast horizon the last three years of the sample (our first forecast is for 2008Q2, with the nowcasting and forecasting models estimated over 1999Q3-2008Q1 and 1999Q3-2007Q4, respectively). This forecast horizon covers the very volatile period of the global economic recession in 2008-2009, thus posing a 'real' challenge for our models (in particular, it will be interesting to see how quickly the models adjust during turning points). As main measure for forecast accuracy, we examine root mean squared errors (RMSE), relative to the AR model. The RMSE's of all forecasting models, relative to the AR benchmark, are summarized in table A in the appendix. In addition, we ran forecast efficiency regressions in the spirit of Mincer and Zarnowitz (1969), whereby actual GDP realizations are regressed on the nowcast or forecast in the past. The main conclusions are in line with

[25] Dynamic factor models also model the time series properties of the factors (see Geweke, 1977; Forni et al., 2000 for more details). Eickmeier and Ziegler (2006) provide a survey of studies conducted with dynamic factor models.

Figure 4: Timeline of the nowcast and forecast



Note: 'AR' refers to the AR benchmark; 'nowcast' and 'forecast' refers to projections with the China economic indicator, indicator model averaging and the static factor model. To keep things simple, we only show one lag for GDP; when estimated, all models can have up to 5 lags.

the information obtained from the relative RMSE's, so to save space, we do not discuss them in detail (the results are reported in table B in the appendix).[26]

Before we continue, it might be worth considering how data manipulation or errors in the data would affect each of the forecasting methodologies. In this context, it is helpful to distinguish between the possibilities that individual series might be deficient and contain 'measurement errors', and data manipulation, notably data smoothing. Errors in individual series could affect the Chinese economic indicator model substantially, as the number of series it uses is relatively small (and the indicator is simply a weighted sum). One outlier could thus change the projection considerably.[27] That said, most of the indicators are not directly related to GDP, and thus local authorities have arguably less incentives to manipulate them. Small errors in series would likely affect indicator model averaging, but large errors in indicators – like structural breaks – would likely simply imply that this particular indicator is no longer useful in predicting current or next quarter GDP, and this particular regression would thus drop out. This suggests that the technique can handle large discrepancies, but is vulnerable to smaller errors. The opposite is the case for the factor

[26]We also conducted Diebold and Mariano (1995) tests, but in most cases the forecast horizon is simply too short to obtain significant differences.

[27]For instance, in mid-2009 Chinese GDP data was questioned, as electricity use – an indicator in the China economic activity indicator – fell sharply, yet GDP only fell from roughly 10 per cent to about 6 per cent. However, if heavy industry was hit harder by the slowdown than less energy-intensive industries, explaining the divergence.

model: As long as the identification of the factors is not affected, the factor model 'sees through' errors in individual indicators. However, large errors could affect identification of the factors, distorting the factor forecast (Boivin and Ng, 2006). Conversely, a smoothing of GDP data would improve accuracy of models with high autoregressive components (to a varying degree, all our models used lagged GDP, except indicator averaging with model 2).

4 Results

4.1 China economic activity indicator

Figure 5 plots the nowcasts and forecasts obtained using the China real activity index, relative to actual GDP.[28] As can be seen, the nowcast performs reasonably well, while the forecast is considerably less accurate. Both nowcast and forecast seem to react relatively slowly to new data, likely due to the presence of lagged GDP in the models.

In terms of prediction accuracy, the real activity indicator outperforms the AR benchmark models, yielding improvements of almost 20 cent for the nowcast in the third month of the quarter, and around 10 per cent when forecasting next quarter's GDP (see table A in the appendix). This illustrates that the China real activity index has its merits in projecting Chinese GDP in real time. That said, while still outperforming the AR benchmark, the China real activity index is less powerful in forecasting next quarter GDP. This suggests that its main merits lie in nowcasting.

4.2 Forecast averaging using indicator models

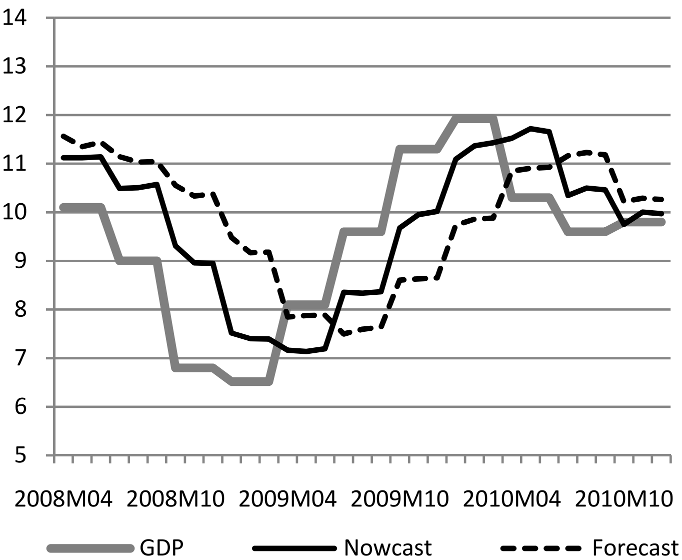
Next, we examine forecast averaging using indicator models. For each of the 33 indicators in our data set, we conduct rolling regressions for the two indicator models 1 and 2. When averaging across models, we drop models where indicators turn out to be insignificant at the 5 per cent level for this particular sample – for instance, if it were the case that freight traffic helps predict GDP in the second and third month of the quarter, but not in the first month, the freight equation would not be included when averaging forecasts for the first month.[29]

This approach yields several interesting pieces of information. Figure 6 shows how often each indicator is used when nowcasting or forecasting GDP. While not providing information on the elasticity of the GDP forecast with regard to individual indicators, the more often an indicator is selected, the more valuable the information is for projecting GDP. As expected, indicators like electricity or steel production, or some of the PMI variables are very valuable, but somewhat unexpectedly, both the Shanghai and the Shenzhen stock market indicators seem

[28] In all charts, GDP, nowcasts and forecasts are all centered at time t (with nowcasts made in the same quarter, while forecasts are made in $t-1$.

[29] To check that we are not overfitting in-sample, we relaxed the constraint that the indicator has to be significant at the 5 per cent level (in fact, we examined different thresholds), but the out-of-sample forecast accuracy deteriorated.

Figure 5: Projecting Chinese real GDP with the China economic activity indicator

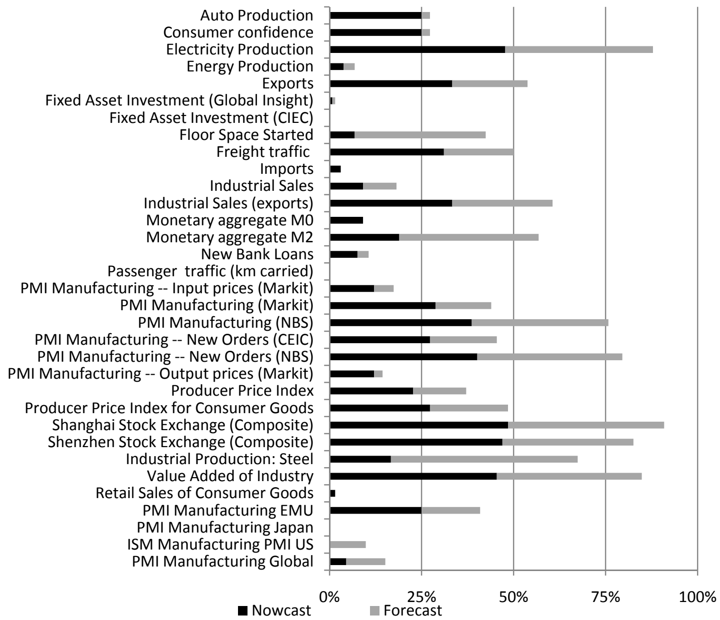


to yield information content, too. In fact, of the 33 indicators in our data set, only 3 are not selected in any of the models (one of which is passenger traffic, a series included in the China activity indicator).

Figure 7 shows the nowcasts and forecasts obtained from indicator models 1 and 2. Both models yield some improvements over the AR benchmarks, in particular for the forecast, where the RMSE, relative to the AR benchmark, drops by almost 40 per cent for model 2 (table A). Note, at the same time, that forecasting performance can be very uneven. Consider, for example, late 2008, when model 1 performed very well, yet model 2 – the model without lagged GDP – was clearly outperformed even by our simple AR benchmark model (the reverse occurred in early 2008). Overall, based on relative RMSE's, we conclude the simple nowcasts of both models are dominated by the China economic activity indicator, but in particular model 2 yields good performance in forecasting.

So far, we have taken simple averages across models, effectively assigning equal weight to all regressions. However, some indicators may be more informative than others, which could translate into differences in nowcasting or forecasting performance. To take this into account, we investigate a number of alternative possibilities to combine the indicator forecasts. First, we consider taking simple averages, but over both models (instead of each model separately), that is, we attach equal weight to all projections made by models 1 and 2. Second, we consider trimmed mean forecasts, whereby we pool over both models, but discard the highest and lowest forecasts each month (the 'outliers'). Third, to explicitly account for differences in model fit, we weigh all models based on

Figure 6: Share of regressions (model 1 and 2) in which individual indicators are significant at the 5 per cent level (in per cent of all regressions)

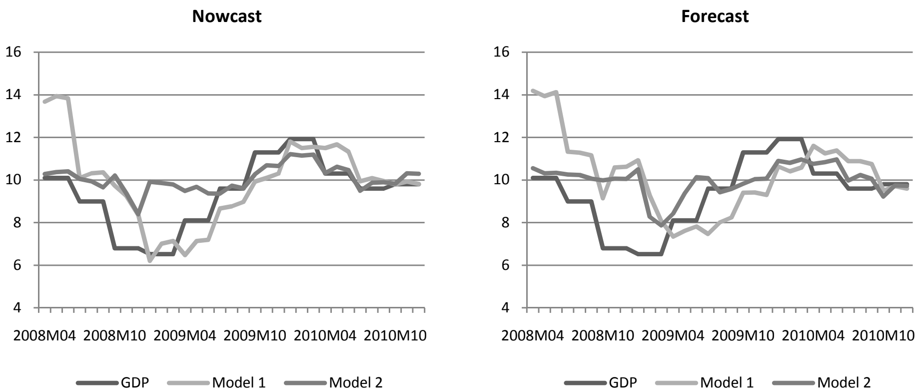


the sum of squared residuals (SSR), calculated at the time of the forecast. As it turns out, in particular the last weighting scheme offers substantial improvements in forecasting accuracy.

The results for the different weighting schemes are shown in figure 8; figure 9 shows relative RMSE's. The performance can be quite impressive. First, under all three weighting schemes, indicator model averaging now outperforms the AR benchmark for all nowcasts and forecasts. As before, forecasts become more accurate during the quarter, reflecting that more information becomes available (that is, RMSE's are lower in the third month of the quarter than in the first month). This suggests that as new information is incorporated, nowcasts and forecasts get closer to the actual value. Second, weighting all indicator models by the sum of squared residuals clearly outperforms the other two weighting schemes. Relative to the AR benchmark, nowcasting accuracy improves by almost 30 per cent in the third month of each quarter, and forecasting accuracy improves by more than 35 per cent. Relative to the China economic activity indicators, SSR-weighted indicator forecasts yield nowcasts that are around 10 per cent more accurate and forecasts that are 25 per cent more accurate in the second and third month.

As an aside, from a practical perspective, indicator model averaging also offers a useful tool to visualize how the forecast evolves, as well as illustrate its uncertainty, by considering a simple histogram of individual forecasts. As an example, consider figure 10, which shows forecasts for 2009Q4, made in

Figure 7: Nowcasting and forecasting Chinese real GDP growth with indicator model averaging



Note: Model 1 contains lags of GDP and indicators at time t; model 2 contains indicators at time t and lagged values of indicators, but no GDP lags.

October, November and December of 2009. Data in October were on the weak side, and the average (unweighted) forecast across all models is 10.2 per cent. In November, the forecast starts to settle in the 10.5-11 per cent range (with an unweighted average of 10.5 per cent), and strong December data increases the forecast to 10.8 per cent, while even pointing to an increasing possibility of growth coming in above 11 per cent (actual GDP growth in 2009Q4 was indeed 11.3 per cent).

4.3 Static factor model

Our last approach is the static factor model. To estimate the model, we need to decide how many factors to include. We normalize all indicators and estimate the factors using principal components. The number of factors retained should be large enough to account for the bulk of common variation in the sample, but small enough to discard factors that represent 'noise' in the data. A helpful metric for the number of factors to retain is the Kaiser-Guttman criterion, which says that factors with an eigenvalue below 1 should be discarded.

On this basis, we find that up to 5 factors are required.[30] Over the full sample, the first five factors explain around 75 per cent of the variation (see figure 11). Inspection of the factor loadings reveals that our first factor seems to load quite heavily on trade-related variables (imports and exports, PMI's for the

[30] A complication is that principal components cannot be estimated when observations are missing. One way to circumvent this issue is to use the 'Expectations Maximization' Algorithm outlined in Stock and Watson (2002). To keep things simple, we estimate the principal factor model over a somewhat more restricted set of variables (17 series in total) to avoid having to generate missing observations. Note that restricting the data set need not be a serious limitation, as Boivin and Ng (2006) show that extending a data set does not necessarily improve the forecasting performance, if additional series are noisy.

Figure 8: Nowcasting and forecasting Chinese real GDP growth with indicator model averaging using different weighting schemes

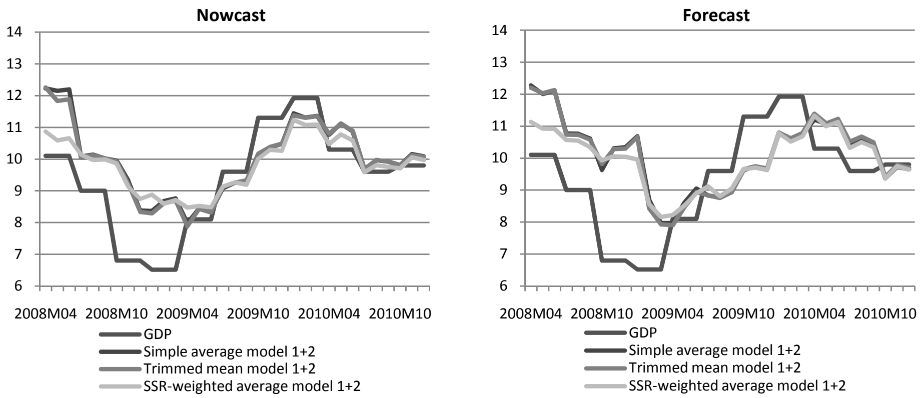


Figure 9: Improvements in projection accuracy using different weighting schemes, relative to an AR benchmark (in per cent)

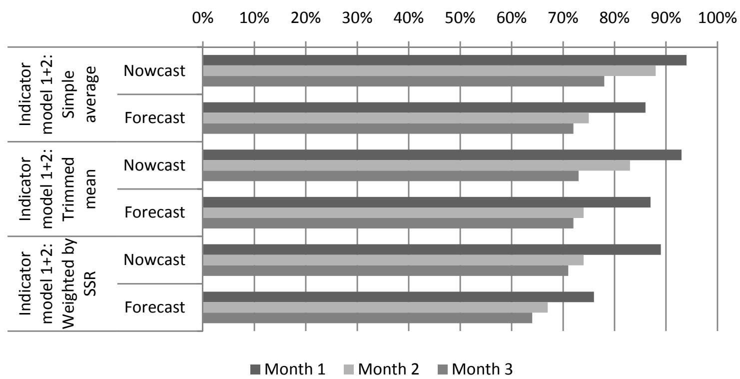
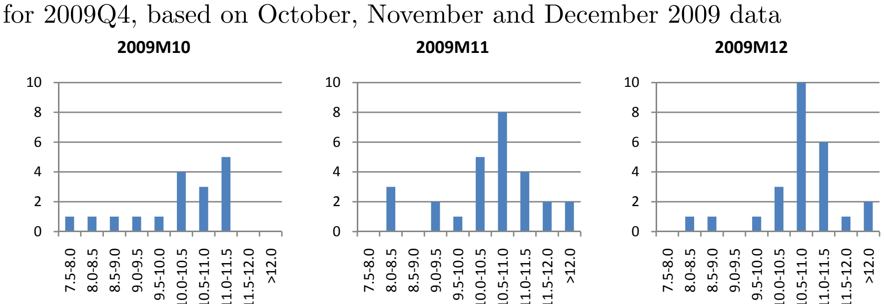
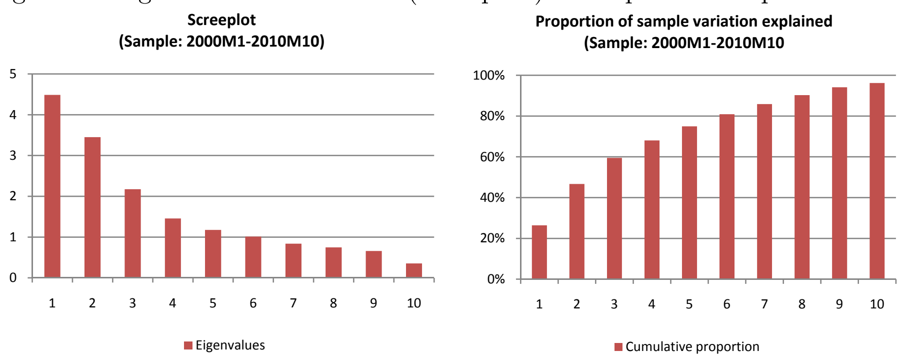


Figure 10: Distribution of current quarter nowcasts (based on simple averaging) for 2009Q4, based on October, November and December 2009 data



Note: Actual Chinese GDP in 2009Q4 was 11.3 per cent.

Figure 11: Eigenvalues of the factors ('screeplot') and explained sample variation



euro area and the United States), while the second factor loads on production-related data such as freight traffic or car production (the interpretation of the other factors is less clear).[31]

The actual projection exercise proceeds in two steps. First, we estimate the first five factors, using data up to the beginning of the forecast horizon. Then, we estimate the nowcasting and forecasting equations. Here, we add one lag of GDP, and in every period select the optimal number of factors based on the Schwarz criterion.

The factor model forecasts and nowcasts are show in figure 12. Overall, the nowcasting and forecasting performance is excellent, and relative RMSE's drop from around 85 per cent in the first month of the quarter to below 50 per cent in the third month of the quarter for both the nowcast and the forecast, compared to the AR benchmark (table A). This makes it the best performing model so far, beating the SSR-weighted indicator model for both the nowcast and the forecast.

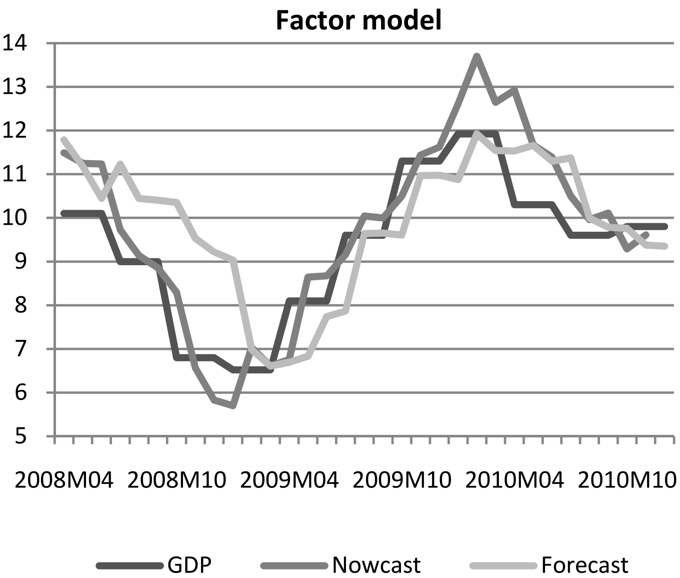
As can be seen, the nowcast is very close to actual values during the downturn in 2008, but overshoots somewhat during the recovery, while an almost reverse pattern in terms of accuracy can be observed for the forecast. Visual inspection also indicates that the factor models seems to be performing best so far in handling the late 2008/early 2009 downturn and subsequent recovery, even if it comes at the cost of overpredicting in early 2010.

5 Averaging across methodologies

So far, we have looked at three different mixed-frequency approaches to project Chinese GDP, and have established that the factor model dominates the other models in terms of nowcasting performance and for most forecasting exercises.

[31]This example has been estimated over the entire sample period; in the forecasting exercises, we re-estimate the factors for every forecast horizon. The loading for the first factor is in line with Mehrotra and Pääkkönen (2011), who also find that the first factor explains around 25 per cent of the variation in the sample.

Figure 12: Nowcasting and forecasting Chinese real GDP growth with the static factor model

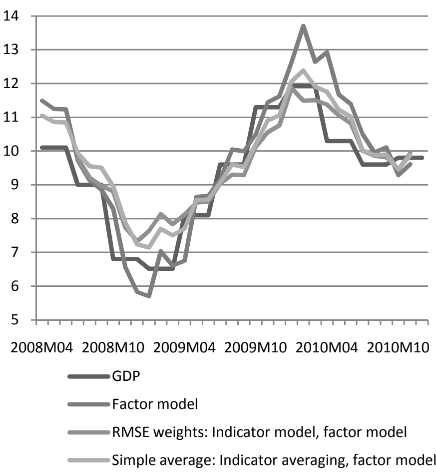


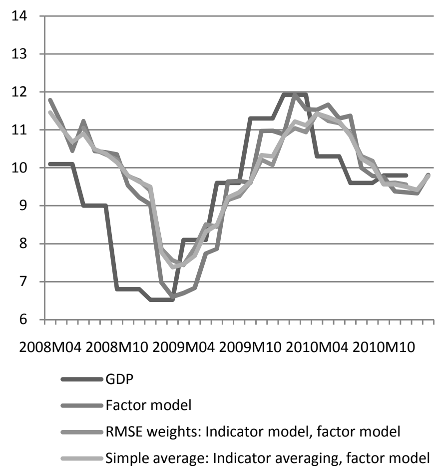
However, the analysis of indicator averaging has also shown that forecast pooling can substantially improve accuracy. In light of this, we next examine whether pooling forecasts across methodologies can further improve accuracy. In short, in most cases the answer is yes.

As was the case with indicator model averaging, different possibilities exist to pool across forecasting methodologies. We report two exercises: simple averages, and weighting forecasting methods according to their RMSE up to the current quarter (i.e. when projecting 2010Q2, we first compute the RMSE's for all models over the forecast horizon up to 2010Q1, then we average across forecasts of each model for 2010Q2, using the 2010Q1 RMSE's as weights). Using these weighting schemes, we evaluate pooled forecasts for the following three models: the China economic activity indicator, the factor model, and the indicator model, whereby individual regressions are weighted by the sum of squared residuals.

Figure 13 presents the nowcast and the forecast for the RMSE-weighted pooled forecast using the factor model and the indicator averaging model, as well as the simple average. Figure 14 shows the relative RMSE's of the pooled forecasts, relative to the AR benchmark. We observe the following. First, regardless of the weighting scheme, the pooled nowcasts and forecasts perform better without the China economic activity indicator. Second, looking at the pooled results without the China economic activity indicator, nowcasting performance of the pooled model improves quite considerably over the factor model. This is particularly the case in the first month, where the relative RMSE's drop from 82 per cent for the factor model to around 65 per cent for models that combine the factor model and the SSR-weighted indicator model averaging. Third, when pooling the factor model and the indicator model, the differences between simple averages and RMSE-based weights are not very large; the Diebold and Mariano (1995) test statistic recommends weighting by RMSE's for the nowcast and simple averages of the two models for the forecast, but the difference is not significant (in fact, in all cases the Diebold and Mariano (1995) test statistic suggests using pooled forecasts, as opposed to relying on one single methodology).

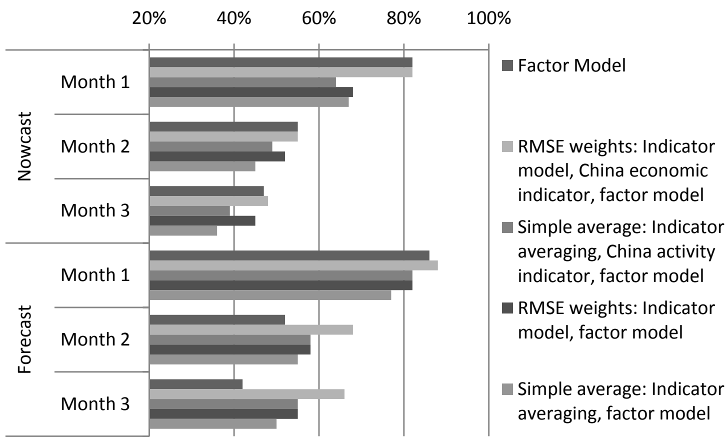
Figure 13: Nowcast and forecast of Chinese real GDP growth when averaging across methodologies





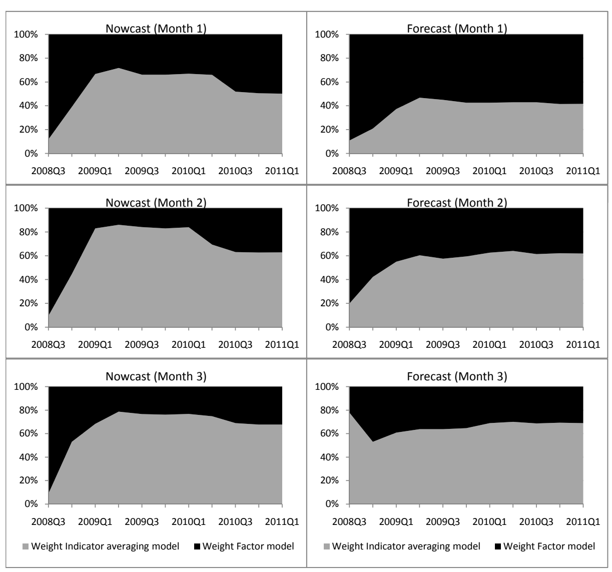
Taken together, a strong case can be made for a combination of indicator model averaging and the static factor model. In combining the two methodologies, the projection performance can be affected by the weights attached to both approaches. To better understand why nowcasting performance increases when switching from simple averaging to an RMSE-based weighted nowcast, while similar improvements are not found for the forecast, we look at the RMSE-based weighting scheme in more detail. Figure 15 shows that all weights change over the forecast horizon, but particularly so for the nowcast. This illustrates that an RMSE-based nowcast implies a relatively larger change in the weights, relative to a simple average, than an RMSE-based forecast, or put differently: a simple average is less close to the RMSE-based weights for the nowcast than for the forecasts. Consequently, taking past performance explicitly into account yields relatively larger improvements for the nowcast than for the forecast (note that in theory, many other weighting schemes are possible, some of which could outperform our RMSE-based approach).

Figure 14: Averaging across methodologies improves accuracy of nowcasts and forecasts of Chinese real GDP growth



Note: Indicator model refers to the model where individual forecasts are weighted by the sum of squared residuals (SSR-weighted average model 1+2).

Figure 15: Weighting schemes for the pooled forecasts



6 Discussion

In this study, we have evaluated different ways to incorporate information from high-frequency indicators to project current and next quarter GDP of the Chinese economy. Using data since 1999Q3 and conducting out-of-sample forecasts for 2008Q2-2010Q4, we evaluate the merits of using the China activity index, developed by Liu et al. (2007), forecast averaging, based on Stock and Watson (2004), and a static factor model, based on Stock and Watson (2002). Comparing our results to a simple AR benchmark, we find that all methods yield substantial improvements in forecasting accuracy. In almost all cases, forecasting accuracy improves as more information becomes available, indicating that all models are able to process the flow of new information efficiently. As regards the accuracy of different methodologies, our results make a strong case that pooling across methodologies can deliver substantially enhanced nowcasts and forecasts. Overall, in line with other studies (Stock and Watson, 2004; Marcellino, 2004), we find that combinations of forecasting methodologies can reduce root mean squared errors up to 60 per cent for the nowcast and the forecast in the third month of the quarter, relative to a naive AR benchmark.

Based on forecasting accuracy, all approaches have their merits, and forecast averaging, in particular, seems a very promising avenue to explore further. However, while high forecasting accuracy is a key element in selecting the best forecasting model, it is not the only one. First, a drawback of using the China activity index for forecasting is that the forecast can only be updated in response to new data summarized in the index, potentially disregarding other sources of information (like stock market movements or information contained in Purchasing Managers' Indices). Similarly, we found that data availability limits the number of series we can use in the static factor model. Forecast averaging based on indicator models is probably the most flexible, as new indicators can easily be added. Consequently, it is possible to evaluate how a given forecast changes in response to, say, the data released this week, providing a much richer picture of the evolution of a forecast during a given quarter. Second, by summarizing information from a broad range of models, forecasting averaging with indicator models provides some safeguard against possible structural breaks. If an indicator ceases to be useful and becomes insignificant in the regression, its corresponding model will automatically be removed and its forecast will no longer be included (conversely, if an indicator becomes useful again, it will automatically be re-included – this provides some flexibility of forecasting during periods of unusually high volatility). Similarly, the static factor model processes information from many series, so errors or breaks in one of them will not immediately affect the projection. To obtain similar robustness to structural breaks, the composition of the China activity index would need to be re-examined frequently. Third, if an indicator is not released in a given month, the indicator model averaging method will simply not provide a projection for this indicator, but as other indicator equations continue to provide useful information, it is still possible to obtain a projection.[32] Lastly, as only a very limited pre-selection of indicators is occurring for forecast averaging and the static forecast, these forecasts are less subject to economist's discretion which indicators to include.

We envision several ways to improve upon our results. Taking averages across methodologies has the additional benefit of providing some insurance against data manipulation. As discussed, forecasting approaches differ in terms of 'vulnerability' to data manipulation. It would be worth exploring this issue in more detail, as well as refining the models we examined by exploring additional weighting schemes (including Bayesian model averaging). We also did not touch upon real-time data issues. Revisions to Chinese data are typically relatively infrequent and comparatively minor (relative to revisions to, say, U.S. or Japanese data), so our forecast evaluation is likely not to be biased by not having historical vintages of data available. However, a more formal investigation could be useful. These are important avenues for future research.

References

Abeysinghe, T. and Rajaguru, G.: 2004, Quarterly real GDP estimates for China and ASEAN4 with a forecast evaluation, *Journal of Forecasting* **23 (6)**, 431447.

Aruoba, S. B., Diebold, F. and Scotti, C.: 2009, Real-time measurement of business conditions, *Journal of Business and Economic Statistics* **27 (4)**, 417–27.

Bailliu, J., Blagrave, P. and Rossiter, J.: 2010, Introducing the Bank of Canada's projection model for the global economy, *Bank of Canada Technical Report* **99**.

Bates, J. and Granger, C.: 1969, The combination of forecasts, *Operational Research Quarterly* **20**, 451–468.

Boivin, J. and Ng, S.: 2006, Are more data always better for factor analysis?, *Journal of Econometrics* **132 (1)**, 169 – 194.

Camacho, M. and Perez-Quiros, G.: 2010, Introducing the EURO-STING: Short-term indicator of euro area growth, *Journal of Applied Econometrics* **25**, 663–694.

Carabenciov, I., Ermolaev, I., Freedman, C., Juillard, M., Kamenik, O., Korshunov, D., Laxton, D. and Laxton, J.: 2008, A small quarterly multi-country projection model with financial-real linkages and oil prices, *IMF Working Paper* **08/280**.

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

[32]In contrast, neither the China economic activity indicator, nor the static factor model, can handle missing observations, unless augmented by, say, the 'Expectations maximization' algorithm (Stock and Watson, 2002).

Diebold, F.: 2007, *Elements of Forecasting*, South-Western College Publishing, Cincinnati.

Diebold, F. and Lopez, J.: 1996, Forecast evaluation and the combination of forecasts, *in* G. Maddala and C. Rao (eds), *Handbook of Statistics*, North Holland, Amsterdam.

Diebold, F. X. and Mariano, R.: 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* **13**, 253–265.

Diron, M.: 2008, Short-term forecasts of euro area real GDP growth: An assessment of real-time performance based on vintage data, *Journal of Forecasting* **27 (4)**, 371–390.

Edge, R. and Gürkaynak, R. S.: 2010, How useful are estimated DSGE model forecasts for central bankers?, *CEPR Discussion Paper* **8158**.

Eickmeier, S. and Ziegler, C.: 2006, How good are dynamic factor models at forecasting output and inflation? A meta-analytic approach, *Deutsche Bundesbank Discussion Paper Series* **1, 42**.

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

Geweke, J.: 1977, The dynamic factor analysis of economic time series, *in* D. J. Aigner and A. S. Goldberger (eds), *Latent Variables in Socio-Economic Models*, North-Holland, Amsterdam.

Goldman Sachs: 2004, Asia-pacific economic analysis, **04/20**, November 9.

Gosselin, M.-A. and Tkacz, G.: 2001, Evaluating factor models: An application to forecasting inflation in Canada, *Bank of Canada Discussion Paper* **18**.

Groen, J. and Kapetanios, G.: 2008, Revisiting useful approaches to data-rich macroeconomic forecasting, *Federal Reserve Bank of New York Staff Reports* **327**.

Hendry, D. and Clements, M.: 2002, Pooling of forecasts, *Econometrics Journal* **5**, 1–26.

Holz, C.: 2005, China governance project: The institutional arrangements for the production of statistics, *OECD Statistics Working Papers* **2005/1**.

Klein, L. R. and Mak, W.: 2005, Initial steps in high-frequency modeling of China, *Business Economics* **40**, 1114.

Lian, M. and Xiaolu, W.: 2000, An evaluation of the reliability of China's statistics on economic growth, *Economic Research (Jingji yanjiu)* **10 (October)**, 3–13.

Liu, L., Zhang, W. and Shek, J.: 2007, A real activity index for mainland China, *Hong Kong Monetary Authority Working Paper* **07/2007**.

Maddison, A.: 1998, *Chinese Economic Performance in the Long Run*, OECD, Paris.

Mak, W.: 2009, Forecasting the sustainability of Chinas economic performance: Early twenty-first century and beyond, *in* L. R. Klein (ed.), *The Making of National Economic Forecasts*, Edward Elgar, Cheltenham, pp. 27–68.

Marcellino, M.: 2004, Forecast pooling for short time series of macroeconomic variables, *Oxford Bulletin of Economic and Statistics* **66**, 91–112.

Mehrotra, A. and Pääkkönen, J.: 2011, Comparing China's GDP statistics with coincident indicators, *Bank of Finland Discussion Papers* **1**.

Mincer, J. and Zarnowitz, V.: 1969, The evaluation of economic forecasts, *in* J. Mincer (ed.), *Economic Forecasts and Expectations*, NBER, New York.

Newbold, P. and Harvey, D.: 2002, Forecast combination and encompassing, *in* M. Clements and D. Hendry (eds), *A companion to economic forecasting*, Blackwell Press, Oxford, pp. 268–283.

OECD: 2008, *OECD system of composite leading indicators*, Organisation for Economic Co-operation and Development, Paris.

Perevalov, N. and Maier, P.: 2010, On the advantages of disaggregated data: Insights from forecasting the U.S. economy in a data-rich environment, *Bank of Canada Working Paper* **2010**.

Rawski, T.: 2001a, China by numbers: How reform affected China's economic statistics, *China Perspectives* **33**, 25–34.

Rawski, T. G.: 2001b, What is happening to china's gdp statistics?, *China Economic Review* **12 (4)**, 347–354.

Sargent, T. and Sims, C.: 1977, Business cycle modeling without pretending to have too much a-priori economic theory, *in* C. Sims (ed.), *New Methods in Business Cycle Research*, Federal Reserve Bank of Minneapolis, Minneapolis.

Schumacher, C. and Breitung, J.: 2006, Real-time forecasting of GDP based on a large factor model with monthly and quarterly data, *Deutsche Bundesbank Discussion Paper Series 1* **33**.

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

Stock, J. and Watson, M.: 2003, Forecasting output and inflation: The role of asset prices, *Journal of Economic Literature* **XLI**, 788–829.

Stock, J. and Watson, M.: 2004, Combination forecasts of output growth in a seven-country data set, *Journal of Forecasting* **23**, 405–430.

The Conference Board: 2001, *Business Cycle Indicators Handbook*, The Conference Board, New York.

Timmermann, A.: 2005, Forecast combinations, *mimeo* .

Yang, Y.: 2004, Combining forecasting procedures: Some theoretical results, *Econometric Theory* **20**, 176–222.

Zheng, J.: 2001, China's official statistics: Growing with full vitality, *China Economic Review* **12 (4)**, 333–337.

A Appendix: Projection accuracy

In what follows, we provide detailed results on forecast accuracy. We use two metrics: first, table A relative root mean squared errors (RMSE's) for all individual projection approaches, as well as forecasts pooled over different methodologies (all RMSE's are relative to the AR benchmark model). Second, table B reports forecast efficiency regressions, following Mincer and Zarnowitz (1969). In these regressions, realized Chinese GDP (y_t) is regressed on nowcasts and forecasts made with approach i at time t and $t-1$.

Formally:

$$y_t = \alpha_i^h + \beta_i^h \hat{y}_{i,t}^h + e_{i,t}^h, \tag{8}$$

Ideally, the coefficients α_i^h and β_i^h should be zero and one, respectively, and the forecast efficiency regressions should have high R^2. The interpretation is as follows: If the intercept differs from zero, the forecast has – on average – been biased; if the slope coefficient is different from one, then the forecast has consistently under or over predicted deviations from the mean. Lastly, a low R^2 signals that GDP is poorly projected.[33]

Overall, relative RMSE's and the forecast efficiency regressions provide qualitatively similar insights: pooling across methodologies helps improve projection accuracy and reduces the bias, in particular when the combined forecast includes the static factor model (pooled forecasts with the static factor model and indicator averaging tend to have slightly lower R^2, but are less biased). In contrast, pooled projections including the China economic activity index tend to perform worse.

[33]When the point estimates of α_i^h and β_i^h are statistically different from zero and one, R^2 is more charitable than relative RMSE's. This is because the forecast makes errors of size $e_{i,t}^h$ only if she knew the values of α_i^h and β_i^h and would adjust $\hat{y}_{i,t}^h$ accordingly (see also Edge and Gürkaynak, 2010).

Table A: RMSE's of all models, relative to the AR benchmark

	Current quarter (Nowcast)			Next quarter(Forecast)		
	Month 1	Month 2	Month 3	Month 1	Month 2	Month 3
HKMA model	87%	82%	81%	95%	90%	90%
Indicator model 1	108%	109%	94%	109%	95%	91%
Indicator model 2	103%	101%	83%	74%	60%	55%
Indicator model 1+2: Simple average	86%	87%	72%	85%	71%	67%
Indicator model 1+2: Trimmed mean	83%	80%	66%	85%	69%	66%
Indicator model 1+2: Weighted by SSR	83%	72%	65%	73%	67%	63%
Factor Model	82%	55%	47%	86%	52%	42%
Simple average: Indicator model (SSR), factor model	67%	45%	36%	77%	55%	50%
Simple average: Indicator model (SSR), factor model, China economic activity indicator	73%	53%	48%	83%	65%	62%
RMSE weights: Indicator model (SSR), factor model	64%	49%	39%	82%	58%	55%
RMSE weights: Indicator model (SSR), factor model, China economic activity indicator	82%	55%	48%	88%	68%	66%

Table B: Forecast efficiency regressions

	Current quarter (Nowcast)			Next quarter(Forecast)		
	Slope	Intercept	R^2	Slope	Intercept	R^2
HKMA model	0.78	1.83	0.48	0.1	8.34	-0.03
Indicator model 1	0.56	3.88	0.49	0.22	7.09	0.02
Indicator model 2	1.86	-9.43	0.52	1.06	-1.16	0.34
Indicator model 1+2: Simple average	1.08	-1.42	0.58	0.61	3.24	0.15
Indicator model 1+2: Trimmed mean	1.14	-1.97	0.63	0.63	3.1	0.17
Indicator model 1+2: Weighted by SSR	1.52	-5.47	0.67	0.85	1	0.22
Factor Model	0.73	2.26	0.85	0.68	2.61	0.41
Simple average: Indicator model (SSR), factor model	1.04	-0.84	0.84	0.83	1.13	0.36
Simple average: Indicator model (SSR), factor model, China economic activity indicator	1.02	-0.58	0.76	0.65	2.91	0.19
RMSE weights: Indicator model (SSR), factor model	1.24	-2.54	0.84	0.82	1.29	0.31
RMSE weights: Indicator model (SSR), factor model, China economic activity indicator	0.92	0.35	0.74	0.58	3.7	0.14