EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1428 / MARCH 2012

SHORT-TERM FORECASTING OF THE JAPANESE ECONOMY USING FACTOR MODELS

by Claudia Godbout
and Marco J. Lombardi

In 2012 all ECB publications feature a motif taken from the €50 banknote.

Claudia Godbout (Corresponding author)
at Bank of Canada, International Department, 234 Wellington Street, Ottawa, Ontario K1A 0G9, Canada;
e-mail: cgodbout@bankofcanada.ca

Marco J. Lombardi
at European Central Bank, Kaiserstrasse 29, D-60311 Frankfurt am Main, Germany; e-mail: marco.lombardi@ecb.europa.eu

Address
Kaiserstrasse 29, 60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19, 60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Internet
http://www.ecb.europa.eu

Fax
+49 69 1344 6000

ISSN 1725-2806 (online)

This paper can be downloaded without charge from http://www.ecb.europa.eu or from the Social Science Research Network electronic library at http://ssrn.com/abstract_id=2020334.

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html

Abstract

While the usefulness of factor models has been acknowledged over recent years, little attention has been devoted to the forecasting power of these models for the Japanese economy. In this paper, we aim at assessing the relative performance of factor models over different samples, including the recent financial crisis. To do so, we construct factor models to forecast Japanese GDP and its subcomponents, using 38 data series (including daily, monthly and quarterly variables) over the period 1991 to 2010. Overall, we find that factor models perform well at tracking GDP movements and anticipating turning points. For most of the components, we report that factor models yield lower forecasting errors than a simple AR process or an indicator model based on Purchasing Managers' Indicators (PMIs). In line with previous studies, we conclude that the largest improvements in terms of forecasting accuracy are found for more volatile periods, such as the recent financial crisis. However, unlike previous studies, we do not find evident links between the volatility of the components and the relative advantage of using factor models. Finally, we show that adding the PMI index as an independent explanatory variable improves the forecasting properties of the factor models.

Keywords: Japan; Forecasting; Nowcasting; Factor models; Mixed frequency
Bank topics: Econometric and statistical methods; International topics
JEL codes: C50, C53, E37, E47

Non-technical summary

Over recent years, factor models have proven useful forecasting tools for dealing with large datasets. Forecasting Japanese GDP and its components is challenging, as it also involves dealing with very volatile data. Resorting to large databases could in principle help to single out common patterns (the 'factors') from multiple data series. To this extent, it is surprising to notice that while the usefulness of factor models in obtaining short-term forecasts for many advanced and developing economies has been documented extensively, little attention has been devoted to the Japanese economy. To the best of our knowledge, no mixed-frequency factor model to forecast the Japanese GDP and its components has appeared in the literature. We see this as a shortcoming, given that policy-makers have a clear interest in providing forecasts not only for monthly series (e.g. inflation or industrial production), but also for (quarterly) GDP and its subcomponents. Factor models provide a simple and convenient way to accomplish this task.

In this paper we explore the usefulness of factor models for forecasting real activity in Japan by resorting to different specifications. More specifically, we construct forecasts of past-, current- and next-quarter GDP, as well as its subcomponents, by using information available on the first, the second and the third month of the quarter. We also assess the performance of factor models over a simple AR specification, as well as a tougher benchmark, an indicator model based on Purchasing Managers index (PMIs).

Overall, we find that factor models perform well at tracking GDP movements and anticipating turning points. For most of the components, we conclude that factor models yield lower forecasting errors than either a simple AR process or an indicator model using PMIs. We do not find evident links between the relative advantage of factor models over the AR process and the volatility of the components to forecast. However, in line with previous studies, we find that the relative improvement from using factor models over a simple AR benchmark is greater over periods of high volatility. For instance, we find that the RMSE of the factor models were around 30 percent lower during the recent recession period (2007-2009) compared to the pre-recession period (1991 to 2007).

We also find that using a forecast average of the factor models à la Stock and Watson (2002) and Giannone, Reichlin and Small (2008) can be optimal and reduces the forecasting errors of the backcasting and nowcasting exercises. We also demonstrate that although the PMI index is already included in the dataset on which factors are computed, it proves nevertheless useful to include it as an independent variable in the factor models. This results from the weighting scheme of the variables entering in the factors, as the variables are weighted based on their ability to represent the common dynamics of the data matrix, which does not necessarily coincide with good forecasting properties for the target variable.

1 Introduction

Forecasting Japanese GDP and its components is challenging, as it also involves dealing with very volatile data. Over the past two decades, GDP growth in Japan has been around 50 percent more volatile than in the United States.[1]

Resorting to large databases could in principle help to single out common patterns (the 'factors') from multiple data series, hence reducing the dimensionality of the data and thus the complexity of the analysis. To this extent, it is surprising to notice that while the usefulness of factor models in obtaining short-term forecasts for many advanced and developing economies has been documented extensively (see, among the others, Stock and Watson (2004), Giannone et al. 2005, Barhoumi et al. 2008, Angelini et al., 2008), little attention has been devoted to the Japanese economy. Fukuda and Onodera (2001) construct a single-index dynamic factor model using leading indicators and demonstrate their usefulness in forecasting coincident monthly indicators such as industrial production, thus omitting GDP developments. Jakaitiene and Dees (2009) also provide factor-based short-term forecasts for a set of countries which includes Japan, but again, they limit their analysis to monthly series. Similarly, Shintani (2005) provides forecasts for Japan by investigating the role of nonlinear model specifications, limiting its analysis to forecasting monthly series such as industrial production. In general, to the best of our knowledge, no mixed-frequency factor model to forecast the Japanese GDP and its components has appeared in the literature. We see this as a shortcoming, given that policy-makers have a clear interest in providing forecasts not only for monthly series (e.g. inflation or industrial production), but also for (quarterly) GDP and its subcomponents. Factor models provide a simple and convenient way to accomplish this task.

The aim of our analysis is to explore the usefulness of factor models for forecasting real activity in Japan by resorting to different specifications. More specifically, we construct forecasts of past-, current- and next-quarter GDP, as well as its subcomponents, by using information available on the first, the second and the third month of the quarter. We also assess the performance of factor models over a simple AR specification, as well as a tougher benchmark, an indicator model based on Purchasing Managers index (PMIs). Overall, we find that forecasting accuracy of factor models is generally higher than both the AR specification and the PMI model.

The paper is structured as follows: in the next section, we present the models used in our analysis, namely factor models à la Stock and Watson (2002) and Giannone, Reichlin and Small (2008), the AR and the PMI indicator model. Next, we present a detailed overview of our dataset, including the time structure of the releases and the publications lags. Subsequently, we discuss the main forecasting results of the baseline experiment for three different samples, the

[1]The annualized quarter-on-quarter growth rate of Japanese GDP has a standard deviation of 3.9 over the sample 1991-2010, compared to only 2.5 in the United States.

full sample, the pre-recession and the recession period. We follow by comparing the different models over different criteria such as the root mean square errors (RMSE) and the proportion of direction changes correctly anticipated. We finish by presenting the advantages of averaging factor models forecasts and by adding the PMI indicator as an independent variable to the factor models.

2 The models

In this section we present the models employed in our forecasting exercise. The first obvious (and naive) alternative against which the models' performance is assessed is a plain univariate AR process, estimated separately for each component of the GDP[2]:

$$\Delta y_t = \mu + \sum_{k=1}^{p} \phi_k \Delta y_{t-k} + \varepsilon_t. \tag{1}$$

The AR order p is selected based on the minimization of the Schwarz criterion, with a maximum lag length of 3.

2.1 Forecasting with leading indicators

Admittedly, the AR benchmark is a very naive one, and very few forecasters will advocate its use in practice. A popular alternative among practitioners is the use of bridge regressions, in which (quarterly) GDP developments are explained by using monthly indicators.[3] Consequently, to give factor models a tougher benchmark, we use a bridging equation model with the headline Purchasing managers Index (PMI)[4] as explanatory monthly variable. PMI's are survey-based diffusion indices, with a value over 50 suggesting an expanding economy, while a value below 50 suggests a contracting economy. Over and above their good forecasting properties, PMI indexes have the advantage of being a very timely indicator, as they are released only a few days after the end of each month. Not surprisingly, PMI indicators have recently received considerable attention in the literature (see among others Koening (2002), Godbout and Jacob (2010) and Rossiter (2010)).

More formally, the PMI model can be defined as:

$$\Delta y_t = \mu + \sum_{k=1}^{p} \phi_k \Delta y_{t-k} + \beta PMI_t + \varepsilon_t, \tag{2}$$

[2]Note that we also considered a random walk in levels, where the forecast of the growth rate of GDP components are equal to the their average growth rate over the sample. However, we find that these two alternatives are generally equivalent and we thus chose to report only the results against the AR process in growth rates.

[3]For a detailed overview of bridge models we refer to Golinelli and Parigi (2007) and references therein.

[4]The headline PMI is a composite index of five survey indices, including new orders, output, employment, suppliers' delivery time and stock of purchases.

where PMI_t denotes a 'quarterly' version of the headline PMI index.[5] To construct the quarterly PMI, we use an AR process to forecast the missing months and then take the average of the three months of the quarter. The AR order is selected based on the minimization of the Schwarz criterion, with a maximum lag length of 3.

2.2 Forecasting with factor models

In a nutshell, the idea underlying factor modes is to represent large datasets using a small number of components able to characterize the main features of the data. The use of factor models originated in the finance literature, where researchers are often faced with large cross-section of stocks returns. The CAPM (Sharpe 1964) and APT (Ross 1976) pricing models are based on a factor representation of the data. Factor models have also been used for macroeconomic applications since the seminal contribution of Sargent and Sims (1977).

More formally, in a factor model, a N-dimensional multiple time series X_t is expressed as

$$X_t = \Lambda F_t + e_t, \tag{3}$$

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. Factor models aim at extracting the most relevant components in large cross-sections of data, and hence represent very well the problems policy-makers face when making decisions: having to look at a wide set of indicators of different nature and summarizing the information they contain about the status of the economy.

Equation (3) is not a standard regression model in that the factors are normally unobservable variables. In some cases (e.g. the CAPM model) the researcher makes assumptions and picks the variables that should be the best factors, but under normal circumstances, F_t has to be estimated. This can be accomplished swiftly and 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.

Given the close resemblance of the factor model setup with the one commonly faced in policy-making – i.e. having to monitor wide sets of indicators and datasets – it is somehow surprising that applications making extensive use of factor models have appeared only recently. One explanation could be that only recently interest has grown in the relevance of data-rich environments for policy-making (Boivin and Giannoni (2006)). A large share of the empirical macroeconomic work is indeed concerned with identifying relationships and comovements between different variables of particular relevance; from a time-series perspective, this is mostly accomplished using the VAR framework. However, VAR models suffer from the so-called "curse of dimensionality", meaning that

[5] Note that equation (2) would of course apply for any other indicator model. We also tried a specification based on the all-industry index and industrial production, but the forecasts were less precise, possibly due to the presence of a publication lag of two months.

they cannot handle too large cross-sections of data because the number of parameters involved in the estimation explodes. The common practice is therefore to focus on small-scale models – usually featuring no more than a few variables. This limits the modeler in taking into account all the set of information available to policy makers. To circumvent this problem, VAR models featuring factor-based sub-structures have been advocated by Bernanke, Boivin and Eliasz (2004) in the setting of the evaluation of monetary policy effects; building on their work Bai and Ng (2006) established limiting and convergence results for VAR models augmented with factors (FAVARs).

The relevance of factor structures is not limited to that of policy analysis, but can also be exploited for forecasting purposes. When forming their ideas about the future path of the economy, policymakers base their judgement by looking at a wide set of indicators and try to extract the best 'signal' out of them. In their seminal contribution, Stock and Watson (2002) consider the prediction of a single time series using a large cross-section of predictors. More specifically, the authors complement equation (3) above with an equation describing the evolution of the 'target' variable y_t:

$$\Delta y_{t+1} = \beta' F_t + \gamma(L)\Delta y_t + \epsilon_{t+1}, \tag{4}$$

where $\gamma(L)$ is a polynomial in the lag operator. This in turn implies that, in order to construct h-step-ahead forecasts, the standard approach of constructing a series of one-step-ahead projections from (4) cannot be employed, and h-step-ahead forecasts have to be constructed directly using the following regression:

$$\Delta y_{t+h} = \beta_h' F_t + \gamma_h(L)\Delta y_t + \epsilon_{t+h}, \tag{5}$$

whose coefficients would of course differ from those of (4). Marcellino, Stock and Watson (2003) report that the direct h-step ahead approach of (5) can be preferable over indirect forecasting as it is more robust to model specification errors. In the following, we will refer to this model as 'SW', for ease of notation.

There are two additional issues that need to be addressed. First, the information set available to policy makers, as summarized by the data matrix X_t, could contain series collected at different frequency. To give a basic example, most of the series commonly used for real developments are at monthly frequency, whereas GDP is quarterly. Second, not all series are released on the same date, as publication lags can be considerable. Hence, the matrix X_t will display a so-called ragged edge, i.e. not all component series will have entries at the end of the sample. Because of these issues, the data matrix will have several missing elements, and as such it will not be possible to extract its principal components. In order to overcome this problem, Stock and Watson (2002) propose to use the expectations maximization (EM) algorithm. The algorithm proceeds as follows: first, missing elements are replaced by their expected values, computed according to (3); second, principal components are extracted from the 'enhanced' data matrix. Since expected values depend on the factors and the loadings, the aforementioned procedure has to be iterated until convergence occurs.

Stock and Watson (2002) show that the information contained in the predictors X_t can be consistently summarized by a restricted number of factors, extracted as principal components of the predictors. Furthermore, they also prove that forecasts generated using (5) are efficient, in the sense that they converge to their (unfeasible) counterpart

$$\Delta y_{t+h} = \beta' F_{t+h} + \gamma(L)\Delta y_{t+h-1} + \epsilon_{t+1}. \tag{6}$$

Finally, the authors document that the forecasting performance of the model outperforms that of standard benchmarks.

In the factor model à la Stock and Watson (2002) the evolution of the factors on the time dimension is not explicitly modeled. Giannone, Reichlin and Small (2008) proposed instead to tackle the issue of short-term forecasting by postulating a parametric model for the evolution of the factors, i.e. an AR(p):

$$F_t = \sum_{t=1}^{p} A_p F_{t-p} + u_t, \quad u_t \sim N(0, Q). \tag{7}$$

This model is akin to dynamic factor structures proposed by Forni et al. (2000), but is estimated using likelihood-based rather than frequency-domain methods. In what follows, we will refer to this model as 'GRS'. Given that F_t is unobservable, the introduction of equation (7) transforms the factor model into a (linear and Gaussian) state-space model, which can be dealt with by the Kalman filter. A closed-form likelihood function can be obtained by conditioning on the filtered values and maximizing it yields parameter estimates. Banbura and Modugno (2010) show how to deal with missing elements in this setting by resorting to the EM algorithm. A by-product of the procedure is a series of filtered values $\hat{F}_t$, which can also comprise forecast values. Hence, projections for the target variables can be constructed as

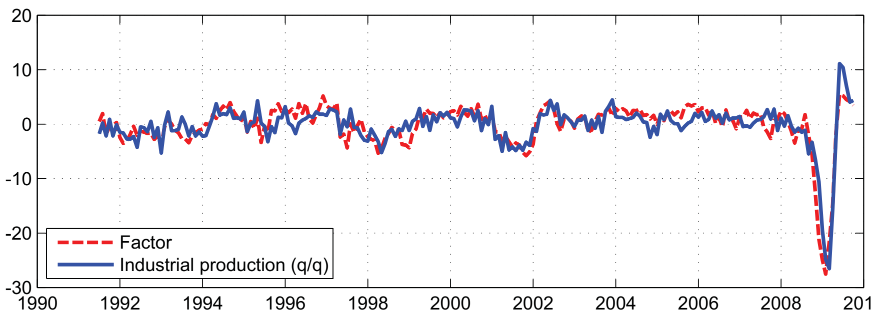
$$\Delta y_{t+h} = \hat{\beta}' \hat{F}_{t+h} + \hat{\gamma}(L)\Delta y_{t+h-1}.$$

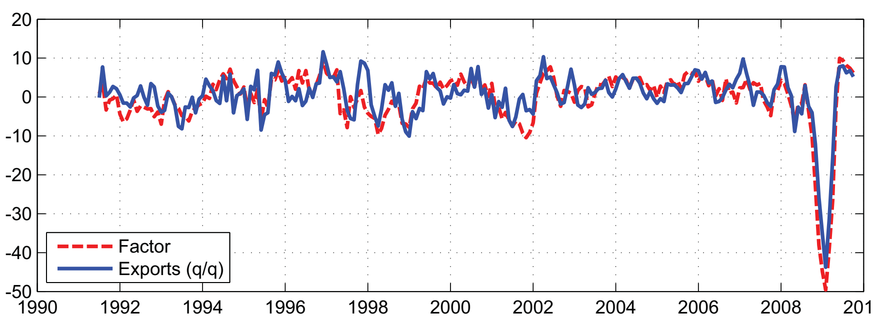
Giannone et al. (2008) as well as the meta-analysis conducted by Eickmeier and Ziegler (2006) suggest that dynamic factor models work better than plain benchmarks, especially when US data are concerned.

3 Data and in-sample model performance

Our factor models employ data for 33 Japanese daily, monthly and quarterly series, covering a broad range of economic concepts, such as output, income, employment, consumer confidence, trade variables, foreign exchange rates, monetary aggregates, interest rates, as well as price and stock market indices; the dataset also includes 4 U.S. monthly variables as well as the U.S. real GDP. When necessary, the series are seasonally adjusted and transformed into a log difference or a level difference, to ensure stationarity. Our objective is to forecast Japanese GDP and its components (exports, imports, capital formation and consumption). As in Giannone et al. (2008), we use pseudo real-time data, in the sense that we consider publication lags when constructing the forecasts.[6] For a complete list of series, transformation operated and publication lags, refer to Table A in the appendix.

Figure 1: Principal component of the data, together with the growth rate of Japan's Industrial Production and Exports (q/q)





Given that the Quandt-Andrews breakpoint test (Andrews 1993) detects the presence of a structural break in Japanese GDP in 1991Q1, we choose to start our analysis in July 1991.[7]

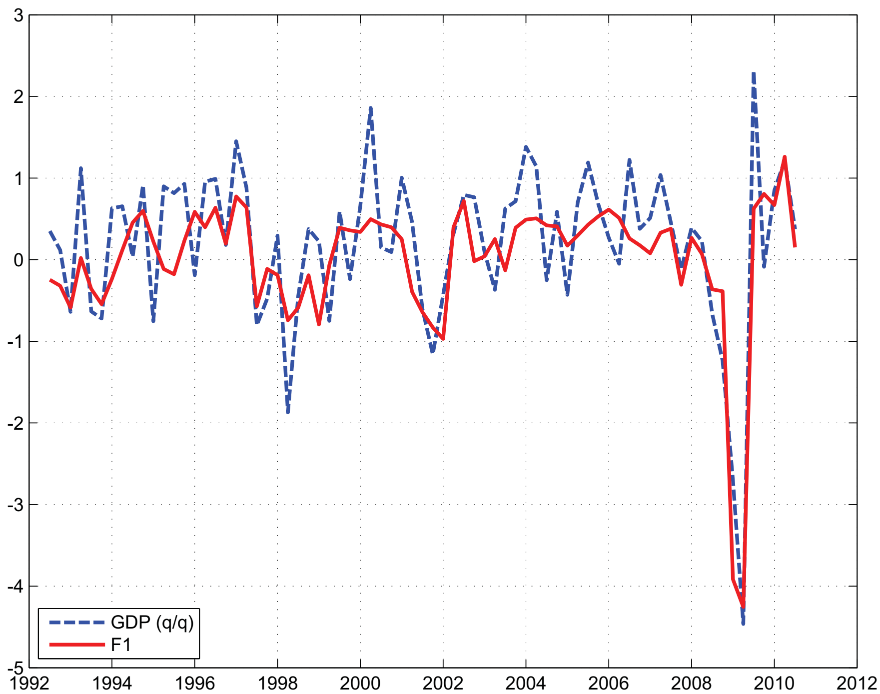
Both the Bai and Ng (2002) and the Alessi et al. (2010) criteria suggest that the optimal number of factors is one.[8] We note that the first factor moves in

[6]Note that we do not use real-time data in this paper. The advantage of using real-time data has not been investigated systematically for the Japanese economy. We leave this aspect for future research. The series used are those available in July 2010

[7]Note that 75 per cent of the series are available at this point in time. We therefore deal with missing values using the EM algorithm.

[8]Note that as a robustness check, we also tried to determine the number of factors based on the out-of-sample RMSE in order to better reflect the out-of-sample nature of our study. This alternative method of selection generally led to confirm the Bai Ng criteria, with the choice of one factor being generally found to reduce the RMSE.

Figure 2: Principal component of the data with real Japanese GDP (q/q)



very close connection with variables such as industrial production and export growth; which is not surprising for an economy heavily reliant on exports (see Figure 1).[9] In Figure 2, we plot the first factor together with real GDP (q/q). We notice that the factor is generally co-moving with GDP, with a correlation between the two series of around 0.8. Moreover, regressing the GDP components on the factor, we find that the first factor explains around 60 percent of the variance of GDP and 80 percent of the variance of exports (see Table 1).

[9]Note that principal components are identified only up to a constant of scale and a rotation matrix, tehreby factors were rescaled in the figures.

Table 1: Proportion of each components' variance explained by the factor.

Δ GDP	0.61
Δ Consumption	0.17
Δ Investment	0.33
Δ Exports	0.79
Δ Imports	0.43

To get a glance at how the models introduced in the previous section capture the dynamics of the data, we estimate them on the full sample, and construct their in-sample fit. In Figure 3, we show the quarterly in-sample fit of factor models for the Japanese GDP (q/q SAAR) over the full sample. On the same chart, we also depict the quarterly fit obtained from a univariate model based on PMIs. At first glance, it appears that all three models display a relatively good performance tracking GDP movements, as well as capturing turning points. Looking at the different models more closely, we can observe that the two factor models are able to better capture the depth and duration of the 2008-2009 recession compared to the PMI model. The two factor models are also able to forecast relatively well the 2001 recession.

Comparing the two factor models, we can notice that although they are strongly comoving, the fit of the GRS model is slightly smoother, probably due to the use of the Kalman filter. Consequently, while the GRS model is able to capture the most relevant trends in GDP, it seems less successful than the SW and the PMI models at precisely tracking the quarterly volatility of GDP. This does not necessarily represent a drawback of the GRS model, since this is likely to lower the volatility of the forecasts and thus reduces the risks of false signals. We verify this assumption later, by looking at the proportion of direction changes correctly anticipated for each models. In fact, depending on the objective of the forecasting exercise, it might be more relevant to get a good sense of the general trend regarding the evolution of the GDP components, as oppose to a volatile signal regarding the specific quarterly numbers. In order to benefit from the advantages of both types of models, we also verify if averaging the forecasts of the two factor models lead to an improvement in the forecasting properties compared to taking each of the models' forecasts independently.

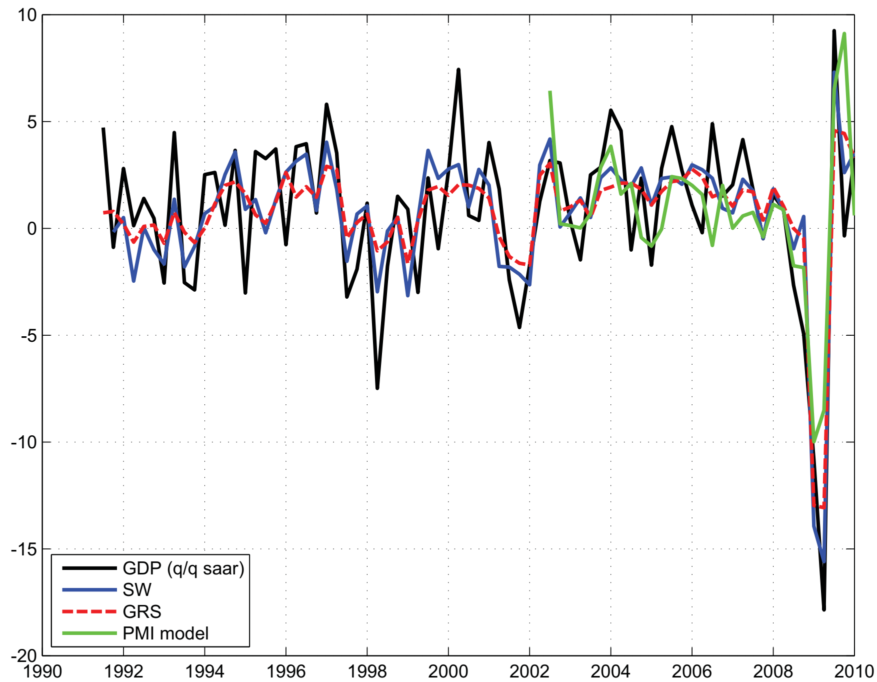
Table 2: Average correlation of series with individual GDP components growth rate over the full sample.

Δ GDP	0.28
Δ Consumption	0.17
Δ Investment	0.18
Δ Exports	0.29
Δ Imports	0.14

4 Out-of-sample forecast evaluation

In this section, we present an out-of-sample evaluation of the forecasting performance of the SW and GRS factor models presented above. Our objective is to forecast GDP and its subcomponents, based on the monthly information provided by the panel of auxiliary variables. The exercise is conducted for the

Figure 3: In-sample fit of the PMI, GRS and SW factor models for the Japanese GDP (q/q SAAR)



last (backcast), current (nowcast), and next-quarter horizon (forecast), based on three different information sets, comprising respectively information up to the first, second and the third month of the quarter. In practice however, factor models could give forecasts every time a new indicator gets released. We first observe results obtained from the full sample (1991-2010), with an out-of-sample period of 4 years, namely 2006 to 2010. Given that the presence of the great recession in this sample could be distorting the results, we also repeat the analysis using only the pre-recession period, thus from 1991-2007, with an out-of-sample period from 2004 to 2007. Finally, we assess the performance of the models during the recession period by looking at the out-of-sample forecasts during the 2-year period of 2007-2009. To compare the performance of the different models, we use two types of criteria. First, we compile the out-of-sample root-mean-square errors (RMSE) for each model. This enables us to determine the absolute fit of the model to the data. However, a forecaster could be more interested in forecasting the direction changes of the series, rather than the specific quarterly numbers. For this reason, in Section 6.3, we also calculate the proportion of direction changes correctly forecasted by each model.

Looking at the out-of-sample fit of the recent recession period more closely (Figure 4), we observe that factor models appear to foresee the recession slightly

Table 3: Mean and standard deviation of GDP components growth rate (q/q, SAAR) over the different samples.

	Full sample	Pre-recession period	Recession period
Mean			
Δ GDP	0.85	1.23	0.36
Δ Consumption	1.10	1.27	0.72
Δ Investment	-1.50	-0.52	-1.83
Δ Exports	4.40	5.46	3.02
Δ Imports	2.88	3.81	0.28
Standard deviation			
Δ GDP	3.90	2.81	5.20
Δ Consumption	3.43	3.37	2.94
Δ Investment	10.82	9.64	12.92
Δ Exports	18.79	8.67	30.73
Δ Imports	12.89	8.23	18.58

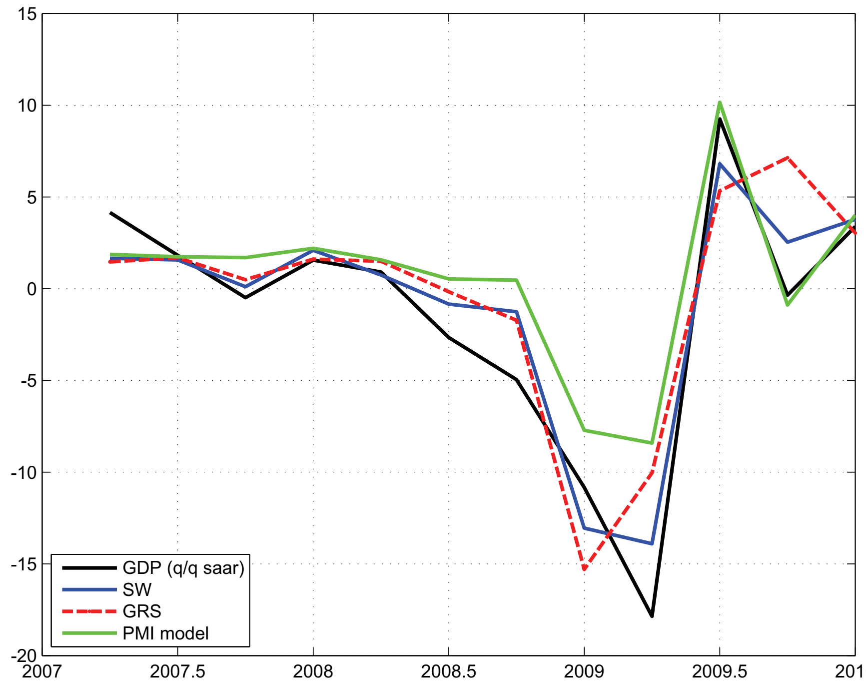
earlier than the PMI indicator model. Moreover, as more data become available, both the SW and GRS factor models seem to outperform the PMI model at representing the depth as well as the duration of the recession. This possibly indicates that while leading indicators can be useful at giving early signals, incorporating more data series might lead to better forecasting properties.

4.1 Forecasting over the full sample

In Table 4, we show the relative RMSE of the PMI, SW and GRS factor models over a simple AR benchmark for the full sample, covering an out-of-sample window of 4 years (2006-2010). A number below one indicates that the model under consideration outperforms the AR benchmark model. One asterisk implies that the difference is significant at the 10 percent level according to the Clark and West test (2007) and two asterisks indicate a significance level of 5 percent. Note also that the second line refers to the number of months before the end of the quarter to forecast, or alternatively in the case of backcasting, the number of months since the end of the quarter to forecast.

Looking at the results, several observations stand out: first, we find that generally speaking, both factor models significantly outperform the AR model for backcasting and nowcasting. On the whole, forecasting errors of factor models for nowcasting and backcasting are found to be between 20 to 60 percent lower than for an AR process. However, apart from the forecast of imports, there do not seem to be advantages of using factor models over an AR process when forecasting the next quarter. This is in line with past literature results highlighting that factor models are generally found to be useful forecasting tools at short horizons, but do not possess great forecasting power at longer horizon

Figure 4: Out-of-sample fit of the PMI, GRS and SW factor models for the Japanese GDP (q/q SAAR) - Backcasting



(Eickmeier and Ziegler (2006)). We also observe that the magnitude of improvement against the AR benchmark generally increases the closer we are from the quarter we aim to forecast. For instance, excluding imports, we find that the errors of the factor models are on average 25 percent lower than those of the AR process in the nowcasting exercise, while they are 32 percent lower for backcasting. Moreover, while nowcasting with the factor models beat the AR process by 20 percent during the first month of the quarter, the improvement reaches around 40 percent by the end of the quarter.

Looking at the components individually, we notice that the factor models possess decent forecasting accuracy for all components as there are no obvious outliers. The largest improvements in terms of forecasting accuracy are found for GDP and exports, with the magnitude of improvement of the SW factor model against the AR reaching 60 per cent for the backcasting and nowcasting of these two components and between 30 and 40 percent for the GRS model. Given that factor models are expected to have the clearest advantage for the more volatile components, it is not surprising to observe a significant advantage of factor models when forecasting exports, which has a high standard deviation (18.8). However, reconciling the large improvements of forecasting accuracy of

Table 4: Relative RMSE of PMI, SW and GRS factor models over an AR process based on information available in the first, second and third month (full sample)

	Backcast		**Nowcast**			**Forecast**		
	+1	+2	-3	-2	-1	-6	-5	-4
				SW				
Y	0.45**	0.37**	1.20	0.75*	0.54**	1.10	1.12	1.10
C	0.75**	0.66**	1.13*	0.75**	0.69**	0.97	1.00	0.94
I	0.69**	0.67**	0.64*	0.70*	0.81*	1.12*	1.14*	0.92**
X	0.69*	0.38**	0.96**	0.64**	0.47*	1.01	0.94*	0.71*
M	0.63	0.45	0.63**	0.71**	0.79**	0.96	0.77**	0.78**
				GRS				
Y	0.61**	0.61**	0.81	0.57*	0.37**	1.12	1.47	1.28
C	0.75**	0.75**	0.85*	0.79**	0.66**	1.16	1.47	1.21
I	1.07	1.06	0.77*	1.14	1.06	0.83**	1.11	0.94**
X	0.73	0.69*	0.64**	0.32**	0.38*	1.02	1.17	0.87
M	1.17	1.17	0.54**	0.50*	0.82**	0.91*	0.87**	0.72**
				PMI				
Y	0.61**	0.61**	1.30	0.92*	0.75*	1.24	1.20	1.18
C	0.72**	0.72**	0.98*	0.82**	0.67**	1.20	1.16	1.08
I	1.00	1.00	1.04	0.98	1.39	1.14	1.15	1.05
X	1.09	1.09	1.29	0.77**	0.61*	1.32	1.25	0.95
M	1.24	1.24	0.82**	0.78**	1.01	1.21	1.21	1.07

Note: A number below one indicates that the model under consideration outperforms the AR benchmark model. One asterisk (*)implies that the difference is significant at the 10 percent level according to the Clark and West (2007) test, and (**) indicates a significance level of 5 percent.

factor models for GDP growth appears surprising at first, given its low standard deviation (3.9). Admittedly, when looking at the correlation of monthly series with the different GDP components (see Table 2), we find that our dataset is largely correlated with GDP and exports, with a respective correlation of 0.28 and 0.29. On the other hand, given the somewhat more limited availability of monthly indicators related to the tertiary sector, our dataset contains relatively fewer information regarding components such as consumption and imports. Not surprisingly, consumption is one of the components with the lowest improvement of the factor model against the AR model. This findings therefore highlights the importance of the choice of the series in the dataset in shaping the forecasts.

In what follows, we assess the performance of the factor models over different samples, enabling us to better define the relative performance of the factor model depending on the sample choice.

4.2 Forecasting the pre-recession period

The first sub-sample to be analyzed is the pre-recession period (1991-2003), with an out-of-sample window of 4 years (2003-2007) for the forecast evaluation.

Given that this period shows less volatility than the full period, this exercise allows us to verify the performance of the factor models in "normal" times. Table 5 shows the relative RMSE of the SW and GRS factor models over the AR benchmark covering the pre-recession period. As we would expect, while the factor models continue to yield lower forecasting errors than the AR process, their relative advantage is generally reduced over this sample, as the volatility of the components is considerably lower (cf. Table 3). For instance, we find that the backcasting and nowcasting errors of the factor models are on average around 20 per cent lower for the factor model than for the AR model over the full period, and only 6 per cent lower in the pre-recession period.

Table 5: Relative RMSE of PMI, SW and GRS factor models over an AR process based on information available in the first, second and third month (pre-recession period)

	Backcast		**Nowcast**			**Forecast**		
	+1	+2	-3	-2	-1	-6	-5	-4
				SW				
Y	0.86**	0.85**	0.96	1.01	0.96*	0.98	1.02	1.07
C	0.98*	1.14	1.00	1.13	1.02	0.98	0.99	1.02
I	0.95	0.89*	1.01	1.06	0.94*	0.90**	0.83**	0.90**
X	0.95*	0.90*	0.82**	0.85**	0.89**	0.82**	0.90**	0.85*
M	0.95	0.90*	0.99	0.96*	0.86*	1.14	1.04	1.01
				GRS				
Y	0.87*	0.86**	0.94	0.99	0.93*	0.97*	0.99	1.01
C	0.96	0.95*	1.00	1.03	0.99*	0.91**	0.92*	1.05
I	0.91	0.90*	1.00	1.05	0.96	0.96**	0.95**	1.00
X	0.95	0.96*	0.78**	0.75**	0.89*	0.84**	0.86**	0.83**
M	0.95**	1.00	0.98	0.97*	0.85**	1.05	1.07	0.92*
				PMI				
Y	0.88*	0.87**	0.94	0.94*	1.00	0.92**	0.91**	1.01
C	1.04*	1.04	1.08	1.07	1.07	0.90**	0.93**	1.10
I	0.95	0.93*	1.11	1.11	1.00	0.97*	0.97*	1.25
X	0.88	0.89	0.86**	0.87**	0.97	0.79**	0.80**	0.83**
M	0.86**	0.87**	0.81**	0.85**	0.84**	1.35	1.33	0.87**

Note: A number below one indicates that the model under consideration outperforms the AR benchmark model. (*) implies that the difference is significant at the 10 percent level according to the Clark and West (2007) test, and (**) indicates a significance level of 5 percent.

4.3 Forecasting the "Great Recession"

Next, we aim at determining the accuracy of the different models during the Great Recession. This period is interesting for several reasons. First and foremost, as the rapidity and depth of the contraction was unprecedented, this period represents an ideal sample for determining the accuracy of the factor models in periods of high volatility, as well as their ability to anticipate the rapid drop in GDP and its components. To assess the performance of the models during the great recession period, we use the full period sample and an out-of-sample window of 2 years, namely from 2007 to 2009. The results are provided in table 5. As expected, during a phase of high volatility, the relative merits of the factor models emerge more clearly. For backcasting and nowcasting, the relative improvement compare to an AR process is around 30 percent larger over this period than for the period excluding the recession (see Figure 5). For instance, the forecasting errors of the backcasting and nowcasting exercises of the SW factor model are respectively 45 per cent and 30 per cent lower than the AR process. For the GRS factor model, forecasting errors are 20 per cent lower than the AR process for the backcasting exercise and 35 percent lower when nowcasting. Moreover, for the SW factor model, the improvement of backcasting and nowcasting over the AR process are significant for all components. For the GRS model, with the exception of nowcasting investment and backcasting imports, the improvement compared to the AR process are also found to be significant for all components. For almost all components, the one-quarter ahead forecast of the SW model is also significantly more precise than the AR process. On the other hand, the forecasting exercise (one-quarter ahead forecast) of the GRS model is yielding several clear misses. For instance, the GDP and consumption forecasting errors are around 30 per cent higher with the GRS model than with the simple AR process.

Overall, Figure 5 shows that the improvements relative to an AR model are considerably larger for the recession period, with an average reduction of forecasting errors of between 30 and 40 percent for that period, compared to an average of 6 percent for the period excluding the recession. Comparing the fit of the factor models to the PMI indicator model, we find that except for the nowcasting of consumption, factor models appear to outperform the PMI model over this sample.

5 Model Comparison

5.1 SW versus GRS

Comparing the fit of the SW and GRS factor models, we find that neither of the models is performing systematically better than the other for all sample and components. In fact, looking at the different sub-samples, we notice that while the forecasting errors of the SW model are generally lower for the full sample as well as the recession period (see Figure 5 and Figures 6 to 8), the GRS factor model performs better for the pre-recession period. This is not surprising, given that the relative stability of the series observed over this period should be to the advantage of the GRS model. Looking at the performance across forecast horizons, we find that the GRS model outperforms the SW model when nowcasting, while the opposite is found when backcasting and forecasting one-quarter ahead. The accuracy of the GRS model is found to be particularly limited when forecasting one-quarter ahead during the recession period, with

Table 6: Relative RMSE of PMI, SW and GRS factor models over an AR process based on information available in the first, second and third month (recession period (2007-2009))

	Backcast		**Nowcast**			**Forecast**		
	+1	+2	-3	-2	-1	-6	-5	-4
				SW				
Y	0.41**	0.35**	0.83	0.63*	0.51**	1.08	0.97	0.91
C	0.69**	0.57**	0.98	0.66**	0.62**	0.93	1.04	0.96
I	0.62*	0.63**	0.54*	0.63*	0.80*	0.82*	0.82*	0.87*
X	0.69*	0.42*	0.85*	0.70*	0.46*	1.10	1.06	0.92
M	0.61	0.43	0.68*	0.83*	0.86	0.97	0.85*	0.84*
				GRS				
Y	0.49**	0.57*	0.67*	0.44*	0.32**	1.14	1.49	1.25
C	0.65**	0.68**	0.81	0.75*	0.59**	1.30	1.73	1.28
I	1.10	0.93*	0.64*	1.13	1.05	0.85*	1.07	0.76**
X	0.73*	0.88	0.71	0.33*	0.38*	1.12	1.36	1.15
M	1.16	1.10	0.61	0.56	0.86*	0.97	0.96*	0.77*
				PMI				
Y	0.56**	0.68*	1.07	0.64*	0.50**	1.14	1.19	1.20
C	0.63**	0.64**	0.94	0.64*	0.49**	1.23	1.24	1.12
I	0.99	0.83*	0.88*	0.90	1.44	1.10	1.07	0.99
X	0.94	1.19	1.16	0.79	0.58*	1.11	1.13	1.13
M	1.08*	0.85*	0.80*	0.84	1.08	0.97*	0.97*	0.90*

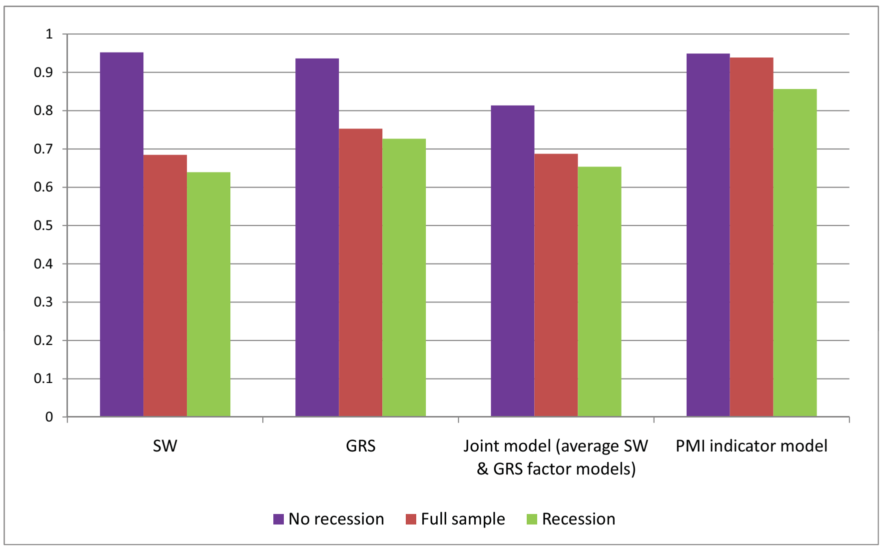
Note: A number below one indicates that the model under consideration outperforms the AR benchmark model. One asterisk (*) implies that the difference is significant at the 10 percent level according to the Clark and West (2007) test, and (**) indicates a significance level of 5 percent.

forecasting errors up to 30 percent larger than those of the AR process. This probably reflects the stable evolution of the GRS forecasts, which in times of high volatility makes it harder to fit the data correctly. For the same reason, we could also expect the relative advantage of the SW factor model to emerge more clearly for the most volatile components. However, the results do not lead us to corroborate this prior. In Figure 9, we show a scatter plot with the standard deviation of each GDP components on the x-axis and the relative RMSE of the SW factor model over the GRS factor model on the y-axis (average of the backcast, nowcast and forecast exercise) for the full sample. As it can be seen, there are no clear links between the volatility of the components and the relative performance of the two types of factor models.

5.2 Factor models versus PMI indicator model

Now that we have determined that the factor models generally outperform the AR process for the backcasting and nowcasting exercises, we can turn to a more difficult benchmark, i.e. the PMI model. Given their timeliness, PMIs have recently received considerable attention as they proved useful indicators during

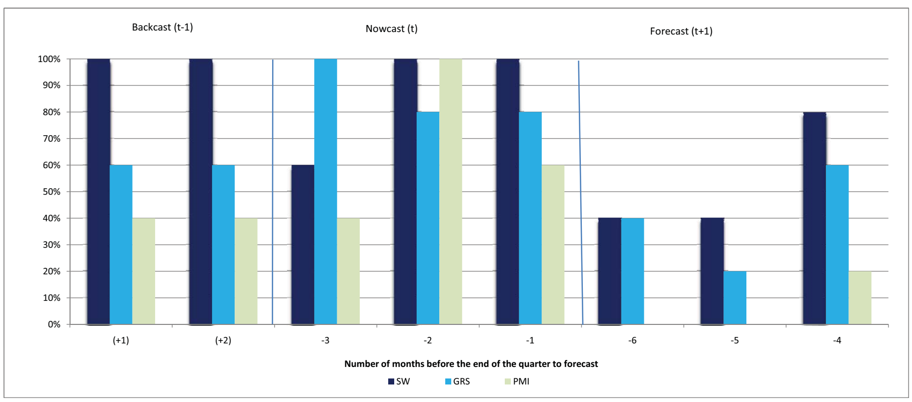
Figure 5: Relative RMSE of the different models over the AR process given different samples (average of all components for the backcasting and nowcasting exercise)



volatile times such as the recession we just encountered. To begin, it is interesting to observe the relationship between the selected factor and the PMI. At first glance, the PMI appears to be relatively well correlated with the factor (see Figure 10), with a correlation of 0.4 between the two series. This is nevertheless lower than the correlation of 0.8 found between the first factor and real GDP growth (q/q SAAR). Moreover, the PMI appears more volatile than both the factor and real GDP growth. Looking at the forecasting results of the PMI indicator model,[10] we observe several elements. First, for all samples, the relative improvement compared to the AR process is lower for the PMI indicator model than for both factor models (see Table 4 to 6 and Figure 6 to 8). This likely reflects the larger information content contained in the factor models. While we could have expected the recession period to be to the advantage of the PMIs (as it was mainly a manufacturing recession), we find that while the forecasting

[10]Note that PMI indictors are only available from 2001. To assure that the results were not affected by the different estimation samples used for the PMI model and the factor models, we re-estimated the factor models over the period post-2001 and did not find significant differences. Most importantly, the relative performance of the different models was not affected by the sample used.

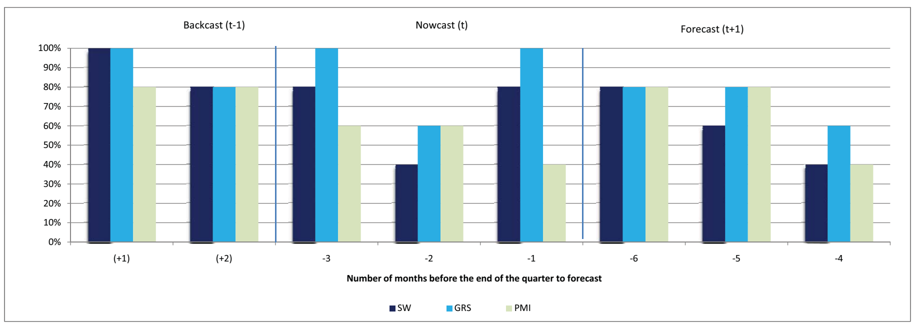
Figure 6: Percentage of times each models beat the AR model over the full sample



errors of the PMI model are smaller than over the full sample (see Figure 5), the PMI model continues to under-perform both factor models over this period. Also, while the PMI model errors are generally smaller than those of the AR process for the backcasting and nowcasting exercises, its one-quarter ahead forecasting performance is much less obvious, with average forecasting errors being 16 percent higher than for the AR process over the full sample (this compares to 8 percent for the GRS factor model and 3 percent lower for the SW factor model). Finally, we observe that the PMI model starts to have decent forecasts only once we have at least one PMI release for the quarter, likely reflecting the very short-term nature of the PMIs. We observe a similar pattern for factor models, but at a smaller extent, possibly a result of the forward looking nature of some variables included in the factors. Annex A presents the fit of each model for all GDP components (out-of-sample). As can be seen, while the PMI appears to be a decent indicator when forecasting GDP and exports, its forecasting power for components such as imports and investment is very limited. Admittedly, this finding is not surprising given the low correlation of the PMI indicator with these components.[11] This highlights one of the advantages of

[11]The correlation of manufacturing PMIs with imports and investment is respectively -0.2 and -0.1, compared to around 0.4 for exports and GDP.

Figure 7: Percentage of times each models beat the AR model over the pre-recession period



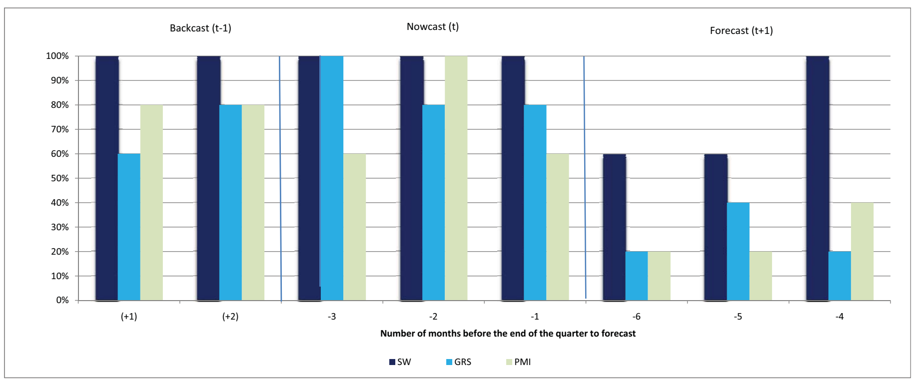
using factor models as this enables the forecaster to use a unique dataset to forecast different components. One other explanation for the poor performance of the PMI indicator model is the possible presence of parameter instability. We verify this aspect in the next section.

5.2.1 Stability of the PMI model

The disappointing performance of the PMI model led us to investigate the stability of the parameters. The PMI index has indeed displayed exceptional gyrations during the recession, which could have seriously compromised the forecasting performance of the model. This is a common risk when using models based on single indicators, which are much more subject to unexpected volatility and/or structural breaks in the chosen indicator than it is the case for factor models, as the estimated factors would somehow smooth and average out excessive volatility in the individual series composing the dataset.

We have therefore conducted a CUSUM stability test (cf. Figure 11) on the coefficients of the PMI model equation for the case of nowcasting (featuring the contemporaneous PMI index) and forecasting (with the PMI index lagged by

Figure 8: Percentage of times each models beat the AR model over the recession period

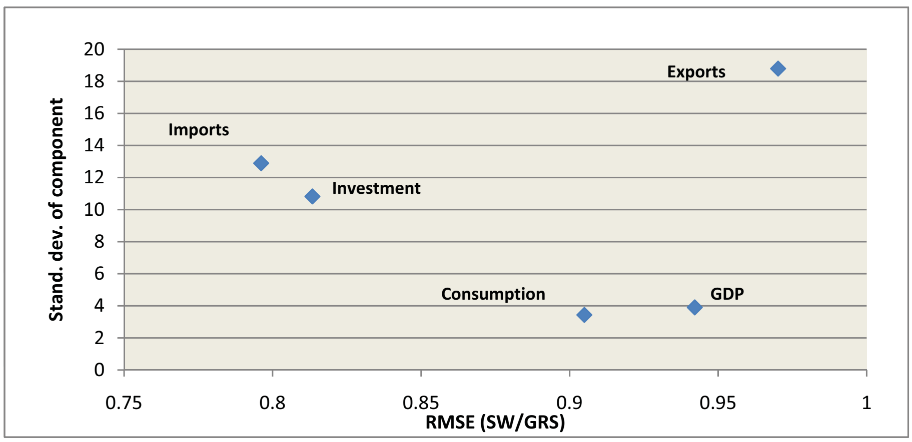


one).[12] In both cases there is evidence of a break in the regression coefficients, but the situation seems much more severe for the case of forecasting. We have also performed a Chow test to evaluate the presence of a breakpoint in 2009Q1[13] and results showed significance at the 5 percent level in the case of nowcasting and at the 1 percent level in the case of forecasting. Looking at the same issue in the setting of the GRS factor model, we find no evidence of structural breaks. This comforts us with the stability of our factor model forecasts. Confirming this result, Banarjee et al. (2006) conclude that one of the advantages of factor models is their performance relative to other standard times series method when dealing with series containing structural breaks.

[12]The case of backcasting has the same specification as the nowcasting.

[13]The date was chosen based on the point at which the PMI index seemed to become more volatile. Due to lack of observations, we could not employ a Quandt-Andrews test to evaluate the most likely date of the break and had to impose it instead.

Figure 9: Relation between Relative RMSE (SW/GRS) and standard deviation of GDP components for full sample

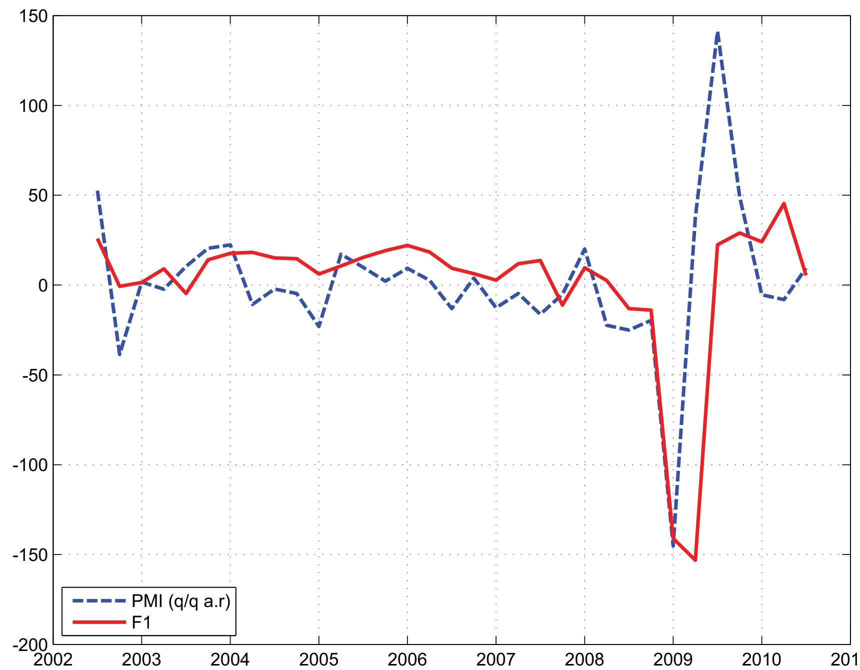


6 Forecast combination

6.1 Combining the SW with the GRS factor model

In previous sections, we have found that neither the SW nor the GRS factor model was performing systematically better for all sample, components or forecast horizon. In fact, we have found that while the stability of the GRS model's forecasts was in time optimal, in other circumstances, the more volatile SW model' forecasts were yielding lower forecasting errors. Consequently, it is possible that an average of the two factor models could be optimal. In Table 7, we thus show the RMSE of this forecast average model (simple average of SW and GRS factor models forecasts) over an AR process for the full sample. As can be seen, it appears that as oppose to both factor models taken individually, this forecast average yields backcasting and nowcasting forecasting errors that are consistently lower than the AR process. On the other hand, there does not seem to be a significant advantage of using the forecast average model when forecasting one-quarter ahead.

Figure 10: First principal components of the data with PMI (q/q SAAR)



6.2 Combining the factor models with the PMI indicator

In the previous section, we have concluded that factor models were generally outperforming the PMI indicator model, probably resulting from the largest set of information included in the factors. However, while this abundance of information clearly represents an advantage of these types of models, it can also represent a possible drawback in a forecasting purpose, due to the relative lack of transparency this can create. Leading indicator models, such as the PMI model are at the opposite end of forecasting models. These models have a small information content, but, on the other hand, possess the advantage of timeliness and transparency. Consequently, for the backcasting and nowcasting exercises, it might be optimal to combine the advantages of both types of models by exploiting the timeliness of the PMIs as well as the large information content of the factor models. In what follows, we therefore try to combine both types of models by adding the PMI indicator as a separate explanatory variable to the factor model.

Equation (5) therefore takes the following form:

$$y_{t+h} = \beta_h' F_t + \gamma_h(L) y_t + \delta PMI_t + \epsilon_{t+h}, \tag{8}$$

Figure 11: Cusum statistic for the parameter stability of the dynamic factor model and PMI model (Dashed lines represent significance of the test at the 5% level.

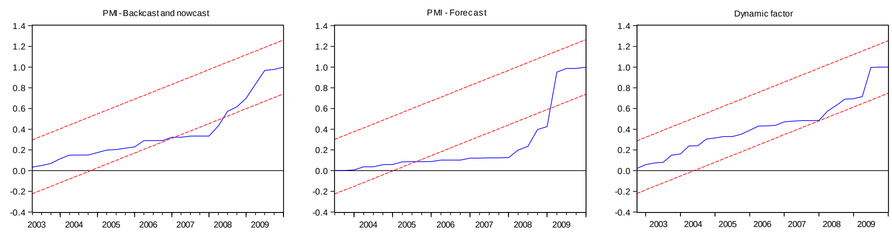


Table 8 shows the relative RMSE of the factor models including the PMI as an explanatory variable against the initial factor models. As it can be seen, the joint model generally outperforms the original factor models, albeit marginally. Not surprisingly, given the timeliness advantage of the PMIs, the largest gains of using a joint model appear for the backcasting and, to a smaller extent, the nowcasting exercise. We can observe that for all components, except imports, using a combined model for backcasting and nowcasting reduces the forecasting errors.

This finding can be surprising at first, given that PMI indicators are already included in the dataset on which factors are computed. However, one need to remember that variables entering the factors are weighted based on their ability to represent the common dynamics of the data matrix, which does not necessarily coincide with good forecasting properties for the target variable. This exercise therefore highlights that giving prominence to leading indicators or other variables known to have interesting forecasting properties can lead to reduced forecasting errors.

Table 7: Relative RMSE of the forecast average model (average SW and GRS factor models) against the AR process (full sample)

	Backcast		**Nowcast**			**Forecast**		
	+1	*+2*	*-3*	*-2*	*-1*	*-6*	*-5*	*-4*
Y	0.52**	0.48**	0.97	0.61*	0.43**	1.09	1.23	1.13
C	0.73**	0.67**	0.96*	0.74**	0.66**	0.99	1.20	1.06
I	0.81*	0.82**	0.70*	0.85*	0.86*	0.88*	0.94*	0.90**
X	0.70*	0.52*	0.77**	0.43**	0.41*	1.00	1.04	0.78*
M	0.87	0.78	0.57**	0.55**	0.78**	0.91	0.79**	0.75**

Note: A number below one indicates that the forecast average outperforms the AR process.

Table 8: Relative RMSE of the SW and GRS factor models including the PMI indicator against the original factor models (full sample)

	Backcast		**Nowcast**			**Forecast**		
	Jan	*Feb*	*Jan*	*Feb*	*Mar*	*Jan*	*Feb*	*Mar*
	SW							
Y	0.78	0.81	0.92	0.82	0.72	1.88	1.10	1.12
C	0.82	0.85	1.15	0.91	1.02	1.49	1.06	1.27
I	0.99	0.93	0.61	0.71	0.58	1.09	1.02	0.98
X	0.38	0.53	0.75	0.83	0.77	2.03	1.07	1.41
M	1.16	1.23	0.77	0.92	0.78	1.46	1.55	1.15
	GRS							
Y	0.99	1.00	0.92	0.92	1.06	1.01	1.03	1.06
C	1.00	1.00	0.97	0.98	0.97	1.00	1.00	1.04
I	0.84	0.86	0.95	0.96	0.87	1.01	1.09	1.31
X	0.65	0.76	1.02	0.98	0.96	0.99	0.96	0.98
M	0.89	0.87	1.06	1.09	1.02	1.02	1.00	1.01

Note: A number below one indicates that the combined model outperforms the factor models without the PMI indicator as an explanatory variable.

6.3 Forecasting the direction of changes

In previous sections, we have compared the different models using the Root-Mean-Squared-Errors (RMSE) criteria. However, policymakers can also be interested in forecasting the direction of the change rather than the specific quarterly number. In this section, we thus compare the models presented above based on the proportion of direction of change correctly forecasted (Table 9)[14]. Looking at the results, several observations stand out. First, as expected, the proportion of changes correctly forecasted increases as we move closer to the quarter we aim to forecast. Second, the factor models outperform the PMI model and the AR process over all horizons. For the backcasting exercise, the factor models correctly anticipate between 75 and 90 percent of direction changes. This is higher than the proportion of changes correctly anticipated by the PMI model (70 percent) and the AR process (55 percent). For the nowcasting exercise, the proportion of direction changes correctly anticipated by the factor models, the PMI indicator model as well as the AR process is largely similar, with a proportion of direction changes correctly anticipated of around 55 percent. Finally, looking at the one-quarter ahead forecasting exercise, we find that while the factor models continue to anticipate correctly more than 50 percent of the direction changes, the PMI model level of success drops to 36 percent in the early months of the quarter. Comparing the two factor models, it appears that the SW model performs better than the GRS model at antic-

[14] Note that the direction changes comparison is done over the full sample, using the same out-of-sample window than the RMSE analysis. The results with a longer out-of-sample window were comparable.

ipating the direction of changes when backcasting, while the reverse is found when nowcasting. This analysis thus confirms the results found in the previous sections and indicates that the factors models are generally able to better anticipate the direction of changes than the PMI and the AR model.

Table 9: Proportion of direction change correctly anticipated by each models (full sample)

	Backcast		**Nowcast**			**Forecast**		
	+1	+2	-3	-2	-1	-6	-5	-4
SW	0.88	0.80	0.50	0.53	0.67	0.50	0.57	0.50
GRS	0.75	0.73	0.53	0.60	0.73	0.57	0.50	0.50
PMI	0.69	0.67	0.53	0.47	0.60	0.36	0.36	0.50
AR	0.56	0.53	0.53	0.53	0.53	0.50	0.50	0.57
Average forecast	0.75	0.80	0.47	0.53	0.73	0.57	0.50	0.57

Note: Joint model is an average of the SW and GRS forecasts.

7 Conclusions

Over recent years, factor models have proven useful forecasting tools for dealing with large datasets. However, while several studies have been conducted on the performance of these models for the United States and the euro area, to our knowledge, these models have not been applied to forecasting Japanese GDP and its components.

In this paper, in order to assess the performance of factor models for forecasting real activity in Japan, we resort to different specifications of both GRS and SW factor models. More specifically, we construct forecasts of last-, current- and next-quarter GDP, as well as its subcomponents, using information available on the first, second and third month of the quarter. We then assess the performance of factor models over a simple AR specification, as well as a tougher benchmark, an indicator model based on PMIs.

Overall, we find that factor models perform well at tracking GDP movements and anticipating turning points. For most of the components, we conclude that factor models (both in SW and GRS version) yield lower forecasting errors than either a simple AR process or an indicator model using PMIs. While previous studies have shown that the advantage of factor models is the clearest for volatile period (D'Agostino et al. 2006) as well as volatile series (Maier and Perevalov 2010), our study confirms the former but not the latter. In fact, we do not find evident links between the relative advantage of factor models over the AR process and the volatility of the components to forecast. However, in line with previous studies, we find that the relative improvement from using factor models over a simple AR benchmark is greater over periods of high volatility. For instance, we find that the RMSE of the factor models were around 30 percent lower during the recent recession period (2007-2009) compared to the pre-recession period (1991 to 2007).

We also find that using a forecast average of the SW and GRS factor model can be optimal and reduces the forecasting errors of the backcasting and nowcasting exercises. We also demonstrate that although the PMI index is already included in the dataset on which factors are computed, it proves nevertheless useful to include it as an independent variable in the factor models. Doing so reduced the forecasting errors by an average of 13 percent for the backcasting exercise and 10 percent for the nowcasting exercise. This results from the weighting scheme of the variables entering in the factors, as the variables are weighted based on their ability to represent the common dynamics of the data matrix, which does not necessarily coincide with good forecasting properties for the target variable. This also underscores the importance of a careful choice of the series when constructing factors, as it may be useful to give prominence to leading indicators such as PMIs, believed to have good forecasting properties. Instead of including some chosen series as additional explanatory variables, as we did in our exercise, a more elegant alternative could be to use a block factor structure, as proposed by Banbura and Modugno (2010), according to which factors are extracted independently from homogeneous blocks of series. We leave this for future research. One additional area for future research could be to assess the performance of factor-augmented error correction model (FECM) in the case of the Japanese economy, following Banerjee et al. (2010). This would have the advantage of incorporating long-run information to the factor models. Finally, it would also be interesting to re-conduct our analysis using real time series, as the large data revisions in Japan could lead to different results.

References

Alessi, L., Barigozzi, M. and Capasso, M.: 2010, Sssimproved penalization for determining the number of factors in approximate factor models, *Statistics & Probability Letters* **80**(23-24), 1806–1813.

Angelini, E., Banbura, M. and Rünstler, G.: 2008, Estimating and forecasting the euro area monthly national accounts from a dynamic factor model, *ECB Working Paper* **953**.

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

Bai, J. and Ng, S.: 2006, Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions, *Econometrica* **74**.

Banbura, M. and Modugno, M.: 2010, Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data, *ECB Working Paper* **1189**.

Banerjee, A. and Marcellino, M.: 2006, Forecasting macroeconomic variables using diffusion indexes in short samples with structural change, *Cambridge University Press* **Chapter 4**, 108–134.

Banerjee, A., Marcellino, M. and Masten, I.: 2010, Forecasting with factor-augmented error correction models, *Department of Economics, University of Birmingham* .

Bernanke, B. S., Boivin, J. and Eliasz, P.: 2004, Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach, *NBER Technical Working Paper* **10220**.

Boivin, J. and Giannoni, M.: 2006, Dsge models in a data-rich environment, *NBER Technical Working Paper* **332**.

Clark, T. E. and West, K. D.: 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of econometrics* **138**, 291–311.

D'Agostino, A., Giannone, D. and Surico, P.: 2006, Unpredictability and macroeconomic stability, *ECB Working Paper* **605**.

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.

Fukuda, S. and Onodera, T.: 2001, A new composite index of coincident economic indicators in japan: How can we improve forecast performances?, *International journal of forecasting* **17**, 483–498.

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.

Godbout, C. and Jacob, J.: 2010, Le pouvoir de prévision des indices PMI, *Bank of Canada Discussion Paper* **3**.

Jakaitiene, A. and Dées, S.: 2009, Forecasting the world economy in the short-term, *ECB Working Paper* **1059**.

Koenig, E. F.: 2002, Using the purchasing managers index to assess the economys strength and the likely direction of monetary policy, *Federal Reserve of Dallas Report* **1**.

Maier, P. and Perevalov, N.: 2010, On the advantages of disaggregated data: insights from forecasting the u.s. economy in a data-rich environment, *Bank of Canada Working Paper* **10**.

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

Parigi, G. and Golinelli, R.: 2007, The use of monthly indicators to forecast quarterly GDP in the short run: An application to the G7 countries, *Journal of Forecasting* **26**(2), 77–94.

Rossiter, J.: 2010, Nowcasting the global economy, *Bank of Canada Discussion Paper* **12**.

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.

Shintani, M.: 2005, Nonlinear forecasting analysis using diffusion indexes: An application to japan, *Journal of Money, Credit and Banking* **37**.

Stock, J. H. and Watson, M. W.: 2002, Macroeconomic forecasting using diffusion indexes, *Journal of Business and Economic Statistics* **20**.

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

Appendix A: In-sample and out-of-sample fit for sub-components of Japanese GDP and Data Description

Figure 12: In-sample fit of the GRS and SW factor models for each Japanese GDP components (q/q SAAR)

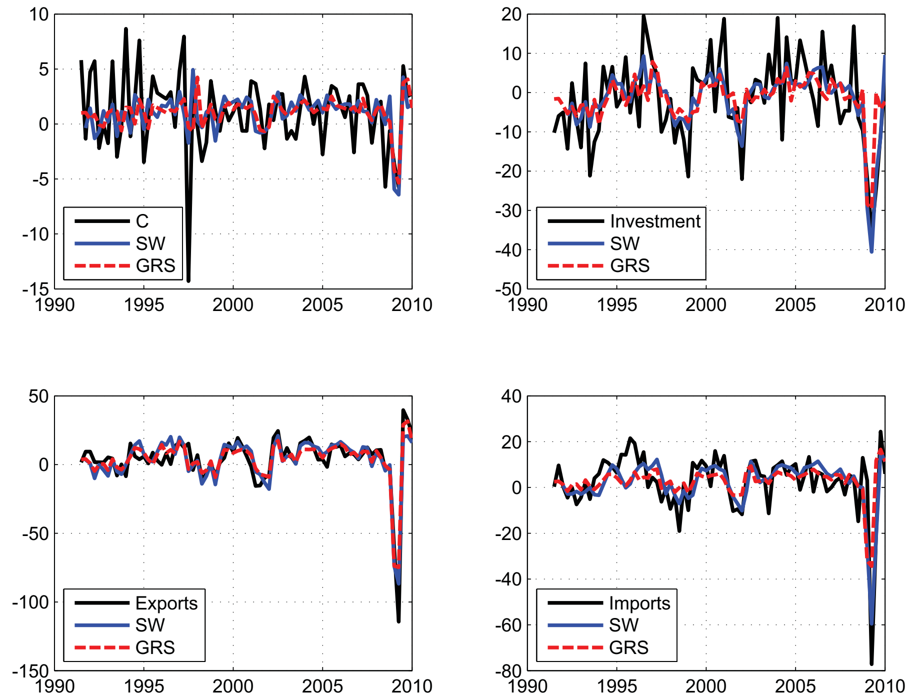


Figure 13: Out-of-sample fit of the GRS, SW and PMI models for each Japanese GDP components (q/q SAAR)

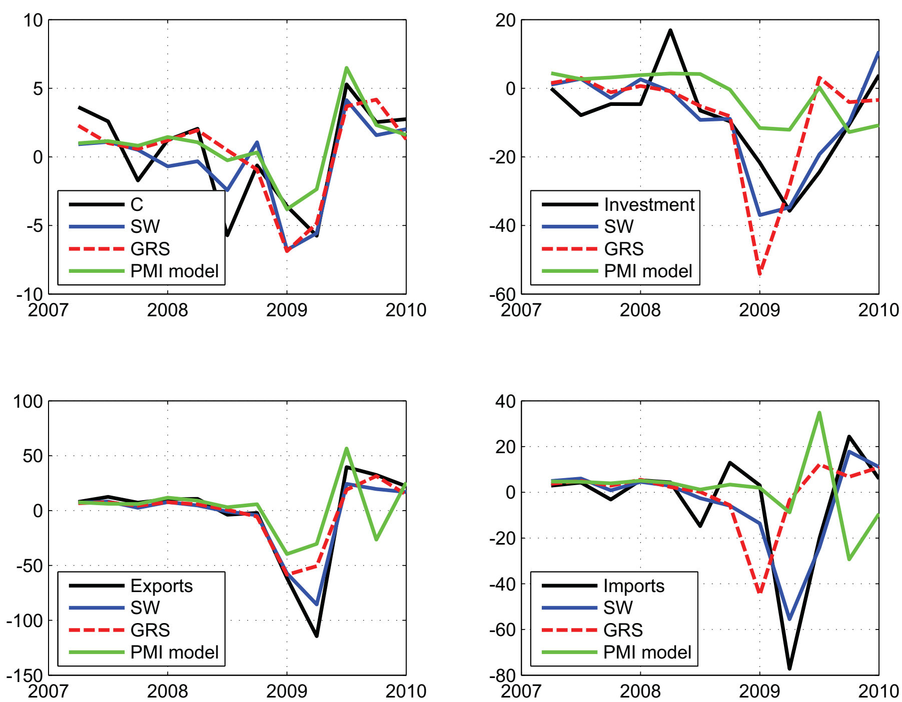


Table 10: Description of variables

Series	Type	Pub. lag	Transformation	SA
Output variables				
Industrial Production	monthly	1	Δ log	Y
Index of operating ratio	monthly	2	Δ log	Y
Business activity index -excluding primary industries	monthly	2	Δ log	Y
Building starts	monthly	2	Δ log	Y
Housing starts	monthly	1	Δ log	Y
Retail sales	monthly	2	Δ log	Y
Living expenditure (average per household)	monthly	2	Δ log	Y
Machinery orders	monthly	2	Δ log	Y
Cabinet Office Consumption index	monthly	2	Δ log	Y
Tankan Survey (manufacturing current)	quarterly	1	Δ level	Y
Tankan Survey (non-manufacturing current)	quarterly	1	Δ level	Y
Employment				
Unemployment rate	monthly	1	Δ log	Y
Ratio of active job openings to active job applicants	monthly	1	Δ log	Y
Confidence and leading indicators				
Manufacturing PMI headline	monthly	0	Δ log	Y
Manufacturing PMI new orders exports	monthly	0	Δ log	Y
Consumer confidence	monthly	1	Δ log	N
Trade and external variables				
Exports (FOB)	monthly	1	Δ log	Y
Imports (CIF)	monthly	1	Δ log	Y
Foreign reserves (end of period)	monthly	2	Δ log	N
Current account balance - factor income	monthly	2	Δ log	Y
Exchange rate (real effective rate)	daily	0	Δ log	N
Money and interest rates				
M1	monthly	2	Δ log	Y
M0	monthly	1	Δ log	Y
Interest rate (uncollateralized overnight rate)	daily	0	Δ log	N
Prices				
Consumer price index	monthly	2	Δ log	Y
Import prices	monthly	1	Δ log	N
Oil price	monthly	0	Δ log	N
Nikkei 225	daily	0	Δ log	N
U.S variables				
U.S. ISM Purchasing managers Index	monthly	1	Δ log	Y
U.S. Industrial Production	monthly	1	Δ log	Y
U.S. fed funds rate	monthly	0	Δ log	N
U.S. CPI total	monthly	1	Δ log	Y
U.S. real GDP	quarterly	2	Δ log	Y
GDP components				
Real consumption	quarterly	2	Δ log	Y
Real private investment	quarterly	2	Δ log	Y
Real Exports good and services	quarterly	2	Δ log	Y
Real Imports good and services	quarterly	2	Δ log	Y
Real GDP	quarterly	2	Δ log	Y