Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Nowcasting Business Cycles: a Bayesian Approach to Dynamic Heterogeneous Factor Models

Antonello D'Agostino, Domenico Giannone, Michele Lenza, and Michele Modugno

2015-066

Please cite this paper as:
D'Agostino, Antonello, Domenico Giannone, Michele Lenza, and Michele Modugno (2015). "Nowcasting Business Cycles: a Bayesian Approach to Dynamic Heterogeneous Factor Models," Finance and Economics Discussion Series 2015-066. Washington: Board of Governors of the Federal Reserve System, http://dx.doi.org/10.17016/FEDS.2015.066.

Nowcasting Business Cycles: a Bayesian Approach to Dynamic Heterogeneous Factor Models

Antonello D'Agostino, European Stability Mechanism
Domenico Giannone, Federal Reserve Bank of New York, CEPR, ECARES, LUISS
Michele Lenza, European Central Bank and ECARES
Michele Modugno, Federal Reserve Board

Abstract

We develop a framework for measuring and monitoring business cycles in real time. Following a long tradition in macroeconometrics, inference is based on a variety of indicators of economic activity, treated as imperfect measures of an underlying index of business cycle conditions. We extend existing approaches by permitting for heterogenous lead-lag patterns of the various indicators along the business cycles. The framework is well suited for high-frequency monitoring of current economic conditions in real time - nowcasting - since inference can be conducted in presence of mixed frequency data and irregular patterns of data availability. Our assessment of the underlying index of business cycle conditions is accurate and more timely than popular alternatives, including the Chicago Fed National Activity Index (CFNAI). A formal real-time forecasting evaluation shows that the framework produces well-calibrated probability nowcasts that resemble the consensus assessment of the Survey of Professional Forecasters.

JEL Classification: C11, C32, C38, E32, E38.

Keywords: Current Economic Conditions, Dynamic Factor Models, Dynamic Heterogeneity, Business Cycles, Real Time, Nowcasting.

The views expressed are those of the authors and do not necessarily reflect those of the European Central Bank, the Eurosystem, the European Stability Mechanism, the Federal Reserve Bank of New York, the Board of Governors of Federal Reserve System. This work was partly supported by the research contracts ARC-AUWB/2010-15/ULB-11 and IAP P7/06 StUDys (DG).

1 Introduction

Macroeconomic and financial variables are characterized by a strong correlation, which is possible only if the bulk of their fluctuations is driven by few common sources. Dynamic factor models (DFM) build on this basic fact to provide a parsimonious and, yet, suitable representation of macroeconomic and financial dynamics. The model assumes that a few unobserved dynamic factors drive the comovement of many observed variables, while the features that are specific to individual series, such as measurement error, are captured by idiosyncratic disturbances.

Dynamic factor models were initially proposed by Geweke (1977) and Sargent and Sims (1977) as a time series extension of the factor models previously developed for cross-sectional data in psychometrics (see Lawley and Maxwell, 1963, for a comprehensive analysis of factor models for serially uncorrelated data). Over the years, factor models have been successfully used in macroeconometrics for structural analysis and forecasting (see Stock and Watson, 2011, for a comprehensive survey). Dynamic factor models have been intensively used in many contexts, ranging from forecasting (Stock and Watson, 2002) and nowcasting (Giannone et al., 2008) to the empirical validation of Dynamic Stochastic General Equilibrium models (Boivin and Giannoni, 2006; Giannone et al., 2006), and provide a reliable statistical framework for the estimation of synthetic indexes of business cycle conditions (Stock and Watson, 1992).

The main feature of business cycle fluctuations is their pervasiveness across the economy.[1] Hence, variables measuring different aspects of the economy can be considered as imperfect measures of a latent common business cycle factor. Formally, the dynamic factor model representation for a set of stationary variables, $x_{i,t}$, $i = 1, ..., n$, is written as follows:

$$x_{i,t} = \sum_{h=0}^{s} \lambda_{i,h} f_{t-h} + e_{i,t} \quad i = 1, ..., n \tag{1}$$

where f_t is the common factor summarizing the state of the economy and $e_{i,t}$, $1 = 1, ..., n$ are the idiosyncratic disturbances. The model is identified by assuming that comovement among variables arises only from a single source, the common factor. This amounts at assuming that $e_{1t}, \ldots, e_{nt}, f_t$ are orthogonal at all leads and lags.

Following Stock and Watson (1992), we will refer to f_t as the *s*ynthetic index of business cycle conditions. Summarizing business cycle condition using a synthetic index rather than observable measures can enhance timeliness and precision. For example, GDP provides a very comprehensive measure of economic activity and summarizes well the business cycle fluctuations, as shown by the fact that recessions roughly correspond to its decline for two consecutive quarters (see Harding and Pagan, 2002). However, GDP is released with a delay, it is subject to revisions and it is characterized by measurement error. Aggregating the information provided by different variables represents a sort of insurance against the measurement error[2] and, in the assessment of the business cycle conditions, it allows to exploit the different sampling

[1] In their pioneering work, Burns and Mitchell (1946, p.3) define the business cycles as the "type of fluctuation found in the aggregate economic activity of nations that organize their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions [...]." Pervasiveness is central also in the definition of the NBER dating committee: "During a recession, a significant decline in economic activity spreads across the economy and can last from a few months to more than a year. Similarly, during an expansion, economic activity rises substantially, spreads across the economy, and usually lasts for several years" (`www.nber.org/cycles/general_statement.html`).

[2] See, for example, the statement of the CEPR dating committee: "To reduce the chance that data revisions might lead the Committee to reconsider its choice of turning points in the future, the Committee examines a wide array of economic data in addition to GDP, such as the individual components of output and labor market data."

frequency and the different timeliness of macroeconomic data releases.[3]

The dynamic factor model is typically specified by assuming $s = 0$ in equation (1).[4] We refer to this set of restrictions as *dynamic homogeneity.* This assumption can represent a straightjacket since it imposes that different indicators have the the same lead-lag pattern along the business cycles and proportional impulse response functions to a common shock, i.e., to an exogenous shock to the synthetic index. For these reasons, inference is typically performed on pre-selected economic indicators that are judged to be coincident along the business cycle, i.e., indicators that "have been tolerably consistent in their timing in relation to business cycle revivals and that at the same time are of sufficiently general interest to warrant some attention by students of current economic conditions" (see Mitchell and Burns, 1938; Moore, 1983).

In this paper, we relax the assumption of dynamic homogeneity and accommodate for heterogeneous dynamics by including a large number of lags ($s >> 0$) in equation (1). The more general structure reduces the risk of model miss-specification, enabling to extract more efficiently the information from economic indicators characterized by a significant degree of dynamic heterogeneity. However, the high level of generality comes at the cost of parameters proliferation. This could increase estimation uncertainty and induce overfitting, which, in turn, could offset the potential benefits of reduced misspecification, and jeopardize the real-time performance of the model. In order to counterbalance these perverse effects, we combine sample information with a prior belief that lagged effects of the common factor are less important the longer the delay.[5]

We conduct inference using data on real GDP and popular US coincident indices of business cycles. Following a fast growing literature on Bayesian factor models, we estimate the full set of posterior densities for the model's parameters and for the unobserved index of business cycle conditions using Montecarlo Markov Chain techniques (MCMC).[6]

Our framework encompasses the traditional approaches to the construction of business cycle indicators. In particular, principal components are proportional to the posterior mode of the unobserved factor associated to a static factor model ($s = 0$ and serially uncorrelated factors f_t and idiosyncratic components $e_{i,t}, i = 1, \ldots, n$) with homogenous signal-to-noise ratio and a flat prior. If the factor loadings are also assumed to be the same ($\lambda_i(L) = \bar{\lambda}$) principal components become simple cross-sectional averages. Allowing for serial correlation, but keeping the model dynamically homogenous ($s = 0$), we obtain the Index of Coincident Economic Indicators of Stock and Watson (1992).

In-sample inference, based on the most recent available data, shows that the factor provides an accurate characterization of the business cycle dynamics in the United States and suggests that dynamic heterogeneity is an important feature of the data. Indeed, the posterior distribution of the common synthetic index provides a more timely account of peaks and troughs when compared with alternative indicators based on dynamic factor models, like the Chicago FED National Activity Index. In addition, the impulse responses of different indicators to a common shock display a relevant degree of heterogeneity.

[3] The dynamic factor model can be cast in a state space form, which provides a natural environment to deal with missing data and mixed frequencies; it is then a suitable tool for the assessment of economic conditions in real time. See Giannone et al. (2008); Aruoba et al. (2009); Camacho and Perez-Quiros (2010); Jungbacker et al. (2011); Bańbura and Modugno (2014), and for surveys Bańbura et al. (2011, 2013).

[4] See, for example, Stock and Watson (1992); Kim and Nelson (1999); Mariano and Murasawa (2003); Aruoba et al. (2009); Camacho and Perez-Quiros (2010); Bańbura et al. (2013).

[5] This is the same logic of the prior beliefs popularized by Litterman (1979) for Bayesian Vector Autoregressions.

[6] See Kim and Nelson (1999); Del Negro (2002); Kose et al. (2003); Justiniano (2004); Bernanke et al. (2005); Del Negro and Otrok (2008); Mackowiak et al. (2009); Moench (2012).

As stressed above, factor models have proved to be successful not only in the extraction of synthetic indicators, but also for nowcasting in real time. We evaluate the accuracy of our model also along this dimension. This is important also because it reveals whether the in-sample properties we just described are genuine features of the data and not only an artifact due to overfitting. More in details, we study the properties of the model-based predictive distributions for GDP growth and compare them with the consensus probability assessments of the Survey of Professional Forecasters (SPF). In order to meaningfully compare the two sets of nowcasts, we take a fully real-time perspective, i.e., we collect the real-time vintages for our variables, which were available at the time the SPF was conducted. Results indicate that the predictive densities are correctly specified - well calibrated - since they cannot be statistically distinguished from the true unconditional data densities. Predictive scores reveal that the predictions of the model are, on average, more accurate than those obtained using a univariate autoregressive benchmark and compare well with the SPF. Overall, the out-of-sample evaluation indicates that dynamic heterogeneity is a genuine and salient feature of the data, and not just the result of overfitting.

The rest of the paper is structured as follows. Section 2 describes the model and the real-time database. Section 3 studies the in-sample properties of our index of business cycle conditions. Section 4 carries out a formal out-of-sample evaluation of the density nowcasts of our model. Section 5 concludes.

2 The model and the database

2.1 The dynamic factor model

We assume that a set of variables $x_{i,t}$, with $i = 1, ..., n$, is characterized by the following equations:[7]

$$x_{i,t} = \lambda_i(L) f_t + e_{i,t}, \quad i = 1, ..., n \tag{2}$$

where $\lambda_i(L) = \lambda_{i,0} + \lambda_{i,1} L + ... + \lambda_{i,s} L^s$. The process for the common factor f_t and the idiosyncratic components $e_{i,t}$, $i = 1, ..., n$ are approximated by finite autoregressive (AR) models:

- $a(L) f_t = u_t, \quad u_t \sim i.i.d.\mathcal{N}(0, 1);$
- $\phi_i(L) e_{i,t} = v_{i,t}, \quad v_{i,t} \sim i.i.d.\mathcal{N}(0, \sigma_i^2), \quad i = 1, ..., n,$

where $a(L) = 1 - a_1 L - ... - a_{p_f} L^{p_f}$ and $\phi_i(L) = 1 - \phi_{i,1} L - ... - \phi_{i,p_e} L^{p_e}$.

The common shocks u_t are assumed to be orthogonal to the idiosyncratic shocks $v_{i,t}, i = 1..., n$, at all leads and lags. In addition, the idiosyncratic shocks are assumed to be mutually orthogonal at all leads and lags. Under this assumption, the model is known as "exact" since it implies that the cross-correlations among observables are only due to the common factor. Although this assumption may be very restrictive, Doz et al. (2012) have shown that the model is robust to

[7] We assume, without loss of generality, that the variables are demeaned and standardized. In practice, we will estimate the model on demeaned and standardized data and we will re-attribute mean and standard deviation after estimation, as it is common practice in the factor model literature.

non Gaussianity and to weak correlation among idiosyncratic components, provided that estimation is carried out with a sufficiently large number of highly collinear variables.

Thanks to the rich dynamics allowed by the polynomials $\lambda_i(L) = \sum_{s=0}^{q} \lambda_{is} L^s$, the model can account for complex heterogeneity in the dynamic effects of the common factors on the observable variables. However, the generality of the model is obtained at the cost of the proliferation in the number of parameters to be estimated. This is the reason why the model is typically estimated with $s = 0$. The most commonly used synthetic indexes are obtained as posterior modes of the following constrained models, when a flat prior is used:

- Principal components: static factor model ($s = p_e = p_f = 0$) with spherical idiosyncratic component, ($\sigma_i^2 = \bar{\sigma}^2$);
- Cross-sectional averages: static factor model ($s = p_e = p_f = 0$) with spherical idiosyncratic component ($\sigma_i^2 = \bar{\sigma}^2$) and homogenous loadings: $\lambda_i(L) = \bar{\lambda}_0$;
- Index of Coincident Economic Indicators of Stock and Watson (1992): Homogenous ($s = 0$) dynamic ($p_e = p_f = 2$) factor model. We will define this model as *DFM*.

The restriction $s = 0$ implies strong homogeneity on the propagation of the common shocks on the variables. In particular, an implication of this assumption is that the fluctuations of all variables are perfectly coincident over the business cycles.[8] We retain the flexibility of the model, relaxing the homogeneity restriction, and we control for the over-fitting due to parameters proliferation by shrinking the model parameters towards those of a simple naïve model, through the imposition of priors. The prior distributions for all the coefficients are centered on zero, with stronger tightness for higher-order lags, so that posterior coefficients of high-order lags of the factors are sufficiently away from zero only if the data strongly favor non-zero values. Formal bayesian inference allows us to combine the information from the data and the prior.

An equivalent representation of the model is obtained by pre-multiplying both sides of equation (2) by $\phi_i(L)$:

$$\begin{aligned} &\phi_i(L)x_{i,t} = \theta_i(L)f_t + v_{i,t}, && v_{i,t} \sim i.i.d.\mathcal{N}(0, \sigma_i^2), \quad i = 1, ..., n, \\ &a(L)f_t = u_t, && u_t \sim i.i.d.\mathcal{N}(0, 1), \end{aligned}$$

where $\theta_i(L) = \phi_i(L)\lambda_i(L)$. Since $\lambda_i(L)$ and $\phi_i(L)$ are unrestricted, we can estimate the model reparameterized in $\theta_i(L)$ and $\phi_i(L)$.[9] The dynamic effects of the common shocks on $x_{i,t}$ can be retrieved by taking the ratio $\lambda_i(L) = \frac{\theta_i(L)}{\phi_i(L)}$. Our priors specified as follows:

- $\sigma_i^2 \sim IG(1, 3)$,
- $\theta_{i,h} \sim N(0, \tau \frac{1}{(h+1)^2})$,
- $\phi_{i,h} \sim N(0, \tau \frac{1}{h^2})$,
- $a_h \sim N(0, \tau \frac{1}{h^2})$.

[8] Dynamic heterogeneity can be taken into account in the context of principal components by including additional new factors, without modeling explicitly that they are lagged versions of each other (see Stock and Watson, 2002). This approach is suitable for forecasting but not for measuring business cycles conditions since it delivers estimated factors that are linear combinations of contemporaneous and lagged values of the index of economic activity.

[9] Quah and Sargent (1993) follow the same strategy.

where h indicates the lag of the factor or the variable to which the coefficient is associated. The prior covariance among coefficients associated to different variables and different lags is set to zero. Notice that the variance of the prior is lower for the coefficients associated with more distant lags. The hyperparameter τ controls the scale of all the variances and effectively governs the overall level of shrinkage. We fix this parameter to the conventional value of 0.2.[10] These priors, including the choice of the degree of overall shrinkage, are similar to those proposed by Litterman (1979) in the context of Bayesian Vector Autoregressive models.[11] Our dynamic factor model with unrestricted dynamics is shortly defined as Heterogenous Dynamic Factor Model (*HDFM*). In order to be able to capture very general dynamics, we specify the model in order to include twelve lags of the observables, the contemporaneous value and twelve lags of the factors in the equations of the observables and twelve lags of the factors in the equations describing the dynamics of the factors.

As stressed in the introduction, we conduct inference using Gibbs sampling techniques. If all data and also the common factor were observed, drawing from the posterior of the parameters is simple since the prior is conjugate. Conditionally on the parameters and the observable data, then the common factors and the missing data can be drawn using simulation smoothers (Carter and Kohn, 1994; de Jong and Shephard, 1995; Durbin and Koopman, 2002).[12] In other words, the Gibbs sampler consists in alternating the following two steps:

- given a draw of the parameters, draw the missing data and the latent factor conditional on the observations using the simulation smoother;
- given a draw of the the full data and the latent factors, draw the parameters from their posterior.

The algorithm is initialized by using the parameters associated to principal components computed by fitting missing data by a spline function.

2.2 Data

We study the in-sample properties of the HDFM model and its accuracy in a real-time forecast evaluation by using a relatively small dataset for the US economy, including the most popular coincident indicators: real GDP (GDP), real disposable income (DSPI), employment (EMP), industrial production (IP), and real retail sales (RRS).[13] In addition, we include the purchasing manager index (PMI) because, due to the timeliness of its release, it provides an extremely useful information.[14]

The variables are transformed in order to achieve stationarity. Real GDP enters in the model in terms of quarter-on-quarter growth rates, while real income, employment, industrial production, and real retail sales enter the model in terms

[10] We leave for future research the task of conducting inference on the degree of prior tightness, which could be done following the lines of Giannone et al. (2012).

[11] We do not need to rescale the variances of the priors to adjust for the different scale of the variables, as it is customary in BVAR applications, since we perform inference using standardized data.

[12] For a general discussion about the formulation of the state space in presence of missing data, see Bańbura et al. (2014).

[13] These are also the most relevant indicators constantly monitored by the NBER to detect and date peaks and throughs in the business cycle.

[14] The importance of survey data for nowcasting has been documented by Giannone et al. (2008), Giannone et al. (2009), Angelini et al. (2011), Lahiri and Monokroussos (2013). For a survey, see Bańbura et al. (2011, 2013).

of month-on-month growth rates. The PMI is stationary by construction, therefore it enters the model without being transformed.[15]

This dataset is characterized by mixed frequencies because real GDP is sampled quarterly, while all the other variables are sampled monthly. In order to deal with this issue, we treat real GDP as observable in the last month of the quarter. The first two months of the quarter are treated as missing observations. This approach is convenient since, as explained in section 2.1, the algorithm used for inference can easily deal with missing data.

The main benchmark in our forecasting evaluation is the Survey of Professional Forecasters (SPF). For the sake of comparability, we exactly replicate the information set available to the professional forecasters at the time they produced their own forecasts.[16] Specifically, the forecasts are generated every quarter with the information available on the 14^{th} of the second month in the quarter, which is roughly in line with the deadline for the submission of the SPF questionnaires. The forecasting evaluation is carried out using eleven years of vintages, ranging from Q1-2003 to Q4-2013. For each real-time data vintage the sample starts in January 1993. We start the evaluation in 2003 in order to have a first estimation sample of ten years. [17]

The real-time exercise introduces an additional source of missing data due to the different availabilities of the data at the time forecasts are generated. In fact, on the 14^{th} of each second month of the quarter, real GDP is available for the previous quarter (e.g., in February real GDP is available up to Q4 of the previous year), employment and PMI are available up to the previous month (e.g., in February they are available up to January), real retail sales, and real disposable income are available up to two months before (e.g., in February they are available up to December). Industrial production is usually released at mid month (between the 13th and the 17th of each month) and, hence, depending on the vintage, it can have either the same availability of employment and PMI or the same availability of disposable income and retail sales.

3 The synthetic business cycle indicator and dynamic heterogeneity

In this Section we perform in-sample inference using data from the last vintage in our dataset (February 2014). The real-time evaluation is conducted in the next Section.

Figure 1 plots the real GDP growth rate against the other variables included in the database.

INSERT FIGURE 1 HERE

Some features stand out. First, all variables tend to comove with GDP, especially during periods of downturn. Second,

[15] Each month, survey respondents are asked to assess their organizations' performance based on a comparison of the current month to the previous month, see http://www.ism.ws/files/ISMReport/ROBBroch08.pdf.

[16] Real time vintages are downloaded from the Federal Reserve Bank of St. Louis http://alfred.stlouisfed.org/.

[17] We use a recursive updating scheme in our out-of-sample forecasting evaluation, i.e., for every vintage the estimation sample always starts in January 1993.

real disposable income, industrial production and real retail sales display very noisy short-run fluctuations, which tend to hide the lower-frequency fluctuations. Third, the variables exhibit a different lead-lag pattern, which is particularly visible around the great recession. PMI and employment growth tend to lag GDP growth; RRS has a more coincident pattern, whereas DSPI and IP display leading dynamics, providing an early signal of the recession.

Figure 2 plots the six variables versus the HDFM business cycle indicator (median, 16th and 84th quantiles of the distribution), which is explicitly devised to account for heterogeneous lead-lag structure of the variables.

INSERT FIGURE 2 HERE

In general, the HDFM indicator tracks our variables very well. This validates the strategy of estimating a dynamic factor model to capture the comovement among the variables. In addition, the indicator is smoother than the individual variables, suggesting that a large part of their high-frequency fluctuations are of idiosyncratic nature. More in details, the indicator is roughly coincident with DSPI, IP and RRS (first three sub-plots) and it clearly leads EMP, PMI and GDP (last three sub-plots), hence it provides an "average" of the variables whose dynamic heterogeneity is properly taken into account. On the other hand, traditional methods for factor extraction would assign most of the weights to the variables with the most persistent dynamics and less volatility (EMP and PMI in our case), and the underlying estimated indicator would be heavily shaped by these series. To illustrate this point, Figure 3 compares the indicator extracted by employing the HDFM in Section 2.1 with four of the most common methods for factor extraction: the average of the monthly variables included in the panel, the first principal component (PC) of the monthly variables, the Chicago Fed National Activity Index (CFNAI) and the factor extracted from a model that imposes dynamic homogeneity. The latter is the posterior mode of the common factor in our model, estimated under the restriction of complete dynamic homogeneity (DFM).[18]

INSERT FIGURE 3 HERE

The HDFM indicator, which is designed to exploit the dynamic heterogeneity of the variables, leads all the other indicators which do not take into account this important feature of the data. It is worth stressing that this is a pure modeling issue, not related at all with the dimension of the information set; indeed, the CFNAI index, which is extracted from a panel of 85 monthly series, also lags the dynamics of our indicator.[19]

These results have also non trivial implications on the traditional simulation exercises, which are performed to assess the system dynamics after some exogenous shock. To clarify this point, Figure 4 reports the impulse response functions (IRFs) of the (log-)levels of the six variables to a common shock, that is an exogenous shock to the synthetic business cycle index. The red line refers to the median IRF estimated by means of the DFM model, which imposes dynamic

[18] Notice that with flat prior the posterior mode of the model parameters corresponds to the Maximum Likelihood estimates. Following Doz et al. (2012) Maximum Likelihood estimation is performed by using the EM algorithm initialized by principal components. The algorithm is modified to account for arbitrary patterns of missing data following the procedure of Bańbura and Modugno (2014). The algorithm has been shown to be computationally efficient and feasible even with high-dimensional data. Recent results by Jungbacker et al. (2011) and Jungbacker and Koopman (2015) show how computational efficiency can be further improved.

[19] See https://www.chicagofed.org/publications/cfnai/index

homogeneity in the effects of the exogenous shock to the synthetic business cycle indicator. The blue lines refer to the IRFs (median, 16th and 84th quantiles of the distribution) of the HDFM. For all variables, the IRFs of the log-levels are obtained by cumulating the IRFs of the growth rates, with the exception of PMI, which is not transformed, and for which the model produces directly the IRFs of the levels.

INSERT FIGURE 4 HERE

The most important difference between the two approaches is that, when dynamic heterogeneity is excluded by assumption, the IRFs have essentially the same dynamics, up to a re-scaling factor given by the factor loadings. Indeed, the cumulated IRFs for a model with dynamic homogeneity is:

$$\frac{\partial x_{i,t+h}}{\partial u_t} = \lambda_{i,0} \sum_{j=0}^{h} b_j \tag{3}$$

where b_j are the coefficients of the polynomial $b(L) = a(L)^{-1}$, and where $a(L)$ is the polynomial that captures the dynamics of the common factor as described in Section 2.1. As noticed above, the only difference among the IRFs of different variables is their loadings $\lambda_{i,0}$. Instead, accounting for the dynamic heterogeneity, the IRFs are allowed to differ:

$$\frac{\partial x_{i,t+h}}{\partial u_t} = \sum_{j=0}^{h} c_{i,j} \tag{4}$$

where $c_{i,j}$ are the coefficients of the polynomial $c(L) = \theta_i(L)\phi_i(L)^{-1}a(L)^{-1}$. In this case, the IRFs for a specific variable will differ not only by a re-scaling factor, but also by the potentially different importance that lags of the variable itself and of the factors have in explaining the fluctuations. This is evident in Figure 4, where the dynamics captured by the red lines are alike among variables. Instead, when a different lead-lag structure among the variables is allowed, the IRFs may have different dynamics. In fact, the blue lines in the figures show that the variables have heterogeneous patterns after the shock. The most striking example is PMI, that displays a clear hump-shaped reaction to a shock to the common component in the HDFM, while this is not the case for the DFM.

4 Evaluation of the density nowcasts

Dynamic factor models are known to perform very well as forecasting tools (see Stock and Watson, 2011, for a survey). However, the specification we advocate in this paper is richly parameterized, hence parameters estimation uncertainty and overfitting are an important concern. For this reason, we evaluate the out-of-sample (real-time) predictive ability of the model.

Bayesian estimation methods allow us to rigorously account for all sources of uncertainty and, hence, we put particular emphasis on density forecast evaluation. We test the density nowcast accuracy of our HDFM against two popular benchmarks: the GDP nowcasts from the Survey of Professional Forecasters and those from a naïve autoregressive model.

To our knowledge, this is the first paper that compares probabilistic forecasts of models and institutions in a fully real-time perspective.

For the sake of comparability, our out-of-sample exercise is designed to replicate the features of the SPF. Specifically, we ask what the model would predict if used in *"real time"* to answer the SPF questionnaire for GDP growth. More in details, we collected 44 vintages of data which were available, in real time, to the forecasters in the quarters between 2003Q1 and 2013Q4 and, at each point in time, we use the model to derive a nowcast of the GDP growth rate in the current calendar year. For example, by using the data vintage available in the first quarter of 2003, we nowcast GDP growth in that quarter and we forecast GDP growth in the subsequent quarters of 2003 to derive the annual growth rate for 2003.

The annual growth rate g_t for GDP in the calendar year t_y is defined as the growth rate in the average level of GDP over the four quarters in year t_y, compared to the average annual level over the four quarters in year $t_y - 1$:

$$g_{t_y} = 100 * \left(\frac{GDP_{Q1,t_y} + GDP_{Q2,t_y} + GDP_{Q3,t_y} + GDP_{Q4,t_y}}{GDP_{Q1,t_y-1} + GDP_{Q2,t_y-1} + GDP_{Q3,t_y-1} + GDP_{Q4,t_y-1}} - 1 \right)$$

where GDP_{Q_j,t_y} is the level of GDP in the j^{th} quarter of year t_y.

Since GDP enters in terms of quarterly growth in the HDFM, the calendar year growth need to be derived starting from the quarterly growth profiles. This is achieved in two steps: first we approximate the year-over-year (yoy) growth rates as a four quarters moving average of annualized quarter-over-quarter (qoq) growth rates; second, we approximate the calendar year growth as the average of the yoy growth rates within the calendar year.[20]

Once again, our choice is driven by the fact that the SPF density forecasts are only publicly available for this definition of the growth rate.[21] The naïve benchmark is an autoregressive model of order two. When looking at data in real time, one issue to address is which data vintage is used to compute the outcome of the target variable. Our choice is to consider the first vintage in the sample in which the data for the full calendar year are made available.

Figure 5 reports the nowcasts (median, 16th and 84th quantiles, dashed lines) of annual GDP growth for the HDFM and the autoregressive model (AR).[22] The blue solid line in the charts refers to the realizations of the annual GDP. For both models, we report the nowcasts computed in each quarter of the year.[23]

INSERT FIGURE 5 HERE

Figure 5 shows that the density nowcasts of the HDFM are generally centered around the outcome, already in the first quarter of the year, differently from the AR nowcasts. The uncertainty on the growth rate of GDP in the calendar year decreases and, by consequence, the nowcast densities become narrower.

[20]Formally, defining t_q as the last quarter of the calendar year of interest, the computation is equivalent to $(1 + L + L^2 + L^3)(1 - L^4) \log GDP_{t_q} \times 400$. For a recent discussion see Richard Crump and Moench (2014).

[21]The SPF also targets the GDP growth rate in the current quarter, but the density forecasts for this definition are not available.

[22]The order of the autoregressive model is set equal to two, as suggested by the Akaike criteria computed in first estimation sample (1993-2003). Results are similar when using only one lag, when the selection is updated recursively and when the coefficients of the benchmark are restricted to those of the random walk.

[23]The density nowcasts of the SPF are not included in the figure because they are available only in terms of histograms.

Next, we evaluate more formally whether the HDFM density nowcasts are a good approximation of the true data densities, by testing the uniformity of the probability integral transforms (PITs).[24] The PITs are the value of the predictive cumulative distribution evaluated at the true realized values of the variables and are widely used to assess the calibration of density forecasts (most recent works include Aastveit et al., 2011; Mitchell and Wallis, 2011; Geweke and Amisano, 2010; Clark, 2011). In fact, Diebold et al. (1998) show that, if the density forecasts approximate well the true density (i.e., are "well calibrated"), then the PITs should be uniformly distributed in the interval $[0-1]$. Assessing the uniformity of the PITs is equivalent to checking whether the inverse normal transformation of the PITs is standard normal. We compare the first fourth sample moments of the PITs inverse normal transformation are different from the first four moments of the standard normal distribution (zero, one, zero and three respectively). Table 2 reports the four sample moments (columns two to five) for each of the nowcasts computed in the four quarters of the year (rows two to five). Following Bai and Ng (2005) we report the heteroskedasticity and autocorrelation consistent (HAC) standard deviation estimator to provide a rough idea of the statistical significance.[25]

Table 2: Tests of normality, HDFM nowcasts.

Quarter	**First moment**	**Second moment**	**Third moment**	**Fourth moment**
Q1	-0.47 (0.58)	0.52 (0.70)	-0.56 (1.02)	0.72 (1.42)
Q2	-0.20 (0.94)	0.83 (0.99)	- 0.09 (1.97)	1.60 (2.50)
Q3	0.35 (0.93)	0.92 (1.22)	0.70 (2.78)	2.20 (5.11)
Q4	0.52 (0.74)	0.77 (0.84)	0.89 (1.24)	1.24 (1.68)

Note: Sample moments in the four quarters. Standard deviation in parentheses.

Table 2 shows that all the sample moments are close to the theoretical values for the standard normal distribution, indicating that the density nowcasts of the HDFM model are well calibrated. We now turn to the analysis of the "relative" accuracy of the HDFM density nowcasts, comparing their log-scores to those of the alternatives.

In the SPF the forecasters are asked to report, among other things, a density forecast by allocating probabilities to ranges of possible future outcomes of the annual growth of GDP. The lower bottom interval and the upper interval of the range are open bins, and the interior bins have equal lengths of 1 percentage points.[26] Individual responses are aggregated by computing average probabilities. For the sake of comparability, we organize the output of the model-based nowcasts (HDFM and AR) along the same lines of the SPF questionnaire. In other words, for all models, we compute the percentage (frequency) of the outcomes that fall in the different bins identified in the Survey of Professional Forecasters.

[24]Notice that our evaluation is more demanding that the traditional residual based diagnostics since the predictive densities are computed in real time, hence accounting for parameter estimation uncertainty and overfitting.

[25]The implied t-statistics should be taken with caution since the asymptotic distribution is non standard due to the recursive estimation of the parameters (see McCracken and Clark, 2013).

[26]Until the first quarter of 2009 the upper bound of the lower bottom interval is 2%. Starting in the second quarter of 2009 it is 3%.

Then, we compute the log-scores, for each model and period, defined as the logarithm of the frequency of the bin including the observed annual GDP growth rate. The higher the mean of the log-scores, the higher the accuracy of the density nowcasts. In Table 3, column one indicates the quarter in which the nowcast for the calendar year is produced. In the second column, we report the average HDFM log-scores. For the AR (column three) and SPF (column four), instead, we report the difference of the average log-scores with the HDFM counterpart: positive values indicate that the average log-score of the specific model is higher than the average log-score of the HDFM for that specific quarter, and viceversa. The HAC estimate of its standard deviation are reported in parentheses.[27]

Table 3: Evaluation of density nowcasts for the calendar year, average log-scores.

Quarter	**HDFM**	**AR minus HDFM**	**SPF minus HDFM**
Q1	-1.27	-0.16 (0.24)	0.15 (0.11)
Q2	-1.17	0.22 (0.13)	0.12 (0.10)
Q3	-0.68	-0.10 (0.12)	-0.03 (0.18)
Q4	-0.18	-0.01 (0.03)	-0.22 (0.07)

Note: HDFM (column two), average log-scores. AR (column three) and SPF (column four), average log-scores minus average HDFM log-scores. Standard deviation in parentheses.

Results in the first columnof Table 3 show that, as expected, the accuracy improves as more information becomes available during the year. The results in the second column indicate that the HDFM is generally more accurate than the AR model. The third column indicates that, while the SPF density nowcasts are more accurate than those of the HDFM in the first two quarters of the year, the opposite is true in the third and fourth quarter of the year. However, the standard deviations of the sample mean of the difference in log-scores (in parentheses) are quite large compared to the average differences in log-scores, so the differences are unlikely to be statistically significant. Since the evaluation sample is short, the forecasting evaluation should not be seen as a horse race, but rather as an assessment of the validity of the model, aiming to ascertain that the accuracy of the density nowcasts is preserved, in spite of the proliferation of parameters resulting from taking into account general patterns of dynamic heterogeneity.[28]

Figures 6 to 8 zoom on three specific calendar years, the 2008, 2009, and 2010. In the three Figures we report the evolution over the four quarters of 2008, 2009, and 2010 of the density nowcasts, in form of histograms, of the HDFM, the AR, and the SPF.

INSERT FIGURE 6 to 8 HERE

[27] See footnote 25.

[28] Results not reported here show that the HDFM model does not significantly outperform the homogenous model in terms of real-time out-of-sample forecasting accuracy. This result indicates that dynamic heterogeneity, although it is a feature of the data, is not so prominent to significantly improve the forecasting performance of the model, at least not in the short evaluation sample considered here.

Figures 6 to 8 show that, particularly in the most acute part of the recession, the HDFM outperforms the AR model and provides very similar outcomes to the SPF. This result shows that accounting for different sources of information is important (for example, surveys were an important source of information to timely capture the great recession). Moreover, it highlights how the HDFM, in spite of its mechanical nature, is able to replicate the outcomes of the survey of professional forecasters that, presumably, incorporates human judgement. These ability of mechanical models to replicate best practice in nowcasting GDP growth has been extensively documented in the context of point forecasts. The finding above indicate that this stylized fact also holds for density forecasts.

5 Conclusions

A synthetic indicator of economic activity should condense, in a timely and reliable manner, the information of several alternative observable measures. The indicator proposed in this paper is based on a dynamic factor model that explicitly allows for dynamic heterogeneity in the effects of the common factor on the variables. Since the model is richly parameterized, we control for overfitting by combining sample information with a prior belief that the effects of lagged factors on the observed economic indicators are more important the shorter the lag. Empirical results support our modeling strategy and indicate that it is important and feasible to account for general patterns of dynamic heterogeneity in the context of dynamic factor models. Indeed, inference based on our framework provides a timely account of the business cycle peaks and throughs and, in real time, it provides accurate and well-calibrated predictive densities.

In this paper, we have focused on a relatively small set of indicators that have been pre-classified as coincident on the basis of a long-established tradition in business cycle analysis. However, the general framework can be used to analyze more general dataset. Indeed, because of the high level of generality, the dynamic heterogenous factor model allows to analyze simultaneously a variety of indicators, without the need of pre-testing or expert judgement for the classification based on lead-lag patterns. This can be particularly important when dealing with datasets characterized by blurred lines of separation between coincident, leading and lagging indicators, as it tends to be the case when considering additional indicators and other countries. Evidence in this direction has been provided by Luciani and Ricci (2013) who have successfully used our methodology to nowcast Norway.

References

Aastveit, K. A., K. R. Gerdrup, A. S. Jore, and L. A. Thorsrud (2011, September). Nowcasting GDP in Real-Time: A Density Combination Approach. Working Papers 0003, Centre for Applied Macro- and Petroleum economics (CAMP), BI Norwegian Business School.

Angelini, E., G. Camba-Méndez, D. Giannone, L. Reichlin, and G. Rünstler (2011, February). Short-term forecasts of euro area GDP growth. *Econometrics Journal 14*(1), C25–C44.

Aruoba, S. B., F. X. Diebold, and C. Scotti (2009). Real-Time Measurement of Business Conditions. *Journal of Business & Economic Statistics 27*(4), 417–427.

Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics 23*(1), 49–60.

Bańbura, M., D. Giannone, and M. Lenza (2014, September). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. Working Paper Series 1733, European Central Bank. International Journal of Forecasting, forthcoming.

Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Nowcasting and the real time data flow. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 2*. Elsevier-North Holland.

Bańbura, M., D. Giannone, and L. Reichlin (2011). Nowcasting. In M. P. Clements and D. F. Hendry (Eds.), *Oxford Handbook on Economic Forecasting*, pp. 63–90. Oxford University Press.

Bańbura, M. and M. Modugno (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics 29*(1), 133–160.

Bernanke, B., J. Boivin, and P. S. Eliasz (2005, January). Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics 120*(1), 387–422.

Boivin, J. and M. Giannoni (2006, December). DSGE Models in a Data-Rich Environment. NBER Technical Working Papers 0332, National Bureau of Economic Research, Inc.

Burns, A. F. and W. C. Mitchell (Eds.) (1946). *Measuring Business Cycles.* NBER Book Series Studies in Business Cycles.

Camacho, M. and G. Perez-Quiros (2010). Introducing the euro-sting: Short term indicator of euro area growth. *Journal of Applied Econometrics 25*(4), 663–694.

Carter, C. and P. Kohn (1994). On gibbs sampling for state space models. *Biometrica 81*, 541–553.

Clark, T. E. (2011). Real-Time Density Forecasts From Bayesian Vector Autoregressions With Stochastic Volatility. *Journal of Business & Economic Statistics 29*(3), 327–341.

de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 2*, 339–350.

Del Negro, M. (2002, March). Asymmetric shocks among U.S. states. *Journal of International Economics 56*(2), 273–297.

Del Negro, M. and C. Otrok (2008). Dynamic factor models with time-varying parameters: measuring changes in international business cycles. Technical report.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review 39*(4), 863–883.

Doz, C., D. Giannone, and L. Reichlin (2012, November). A Quasi-Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models. *The Review of Economics and Statistics 94*(4), 1014–1024.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*(3), 603–616.

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

Geweke, J. and G. Amisano (2010, April). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting 26*(2), 216–230.

Giannone, D., M. Lenza, and G. Primiceri (2012). Priors for vector autoregressions. Discussion paper 8755, C.E.P.R. Discussion Papers. The Review of Economics and Statistics, forthcoming.

Giannone, D., L. Reichlin, and L. Sala (2006, May). VARs, common factors and the empirical validation of equilibrium business cycle models. *Journal of Econometrics 132*(1), 257–279.

Giannone, D., L. Reichlin, and S. Simonelli (2009, October). Nowcasting Euro Area Economic Activity In Real Time: The Role Of Confidence Indicators. *National Institute Economic Review 210*(1), 90–97.

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

Harding, D. and A. Pagan (2002, March). Dissecting the cycle: a methodological investigation. *Journal of Monetary Economics 49*(2), 365–381.

Jungbacker, B., S. Koopman, and M. van der Wel (2011, August). Maximum likelihood estimation for dynamic factor models with missing data. *Journal of Economic Dynamics and Control 35*(8), 1358–1368.

Jungbacker, B. and S. J. Koopman (2015). Likelihood-based dynamic factor analysis for measurement and forecasting. *Econometrics Journal, volume=18, pages=C1-C21*.

Justiniano, A. (2004). Factor models and mcmc methods for the analysis of the sources and transmission of international shocks. Thesis (ph. d.), -Princeton University.

Kim, C.-J. and C. R. Nelson (1999, June). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, Volume 1 of *MIT Press Books*. The MIT Press.

Kose, M. A., C. Otrok, and C. H. Whiteman (2003, September). International Business Cycles: World, Region, and Country-Specific Factors. *American Economic Review 93*(4), 1216–1239.

Lahiri, K. and G. Monokroussos (2013). Nowcasting US GDP: The role of ISM business surveys. *International Journal of Forecasting 29*(4), 644–658.

Lawley, D. N. and A. E. Maxwell (Eds.) (1963). *Factor Analysis as a Statistical Method*. Butterworths.

Litterman, R. (1979). Techniques of forecasting using vector autoregressions. Federal Reserve of Minneapolis Working Paper 115.

Luciani, M. and L. Ricci (2013, February). Nowcasting Norway. Working Papers ECARES ECARES 2013-10, ULB – Universite Libre de Bruxelles. *International Journal of Central Banking*, forthcoming.

Mackowiak, B., E. Moench, and M. Wiederholt (2009). Sectoral price data and models of price settingv. *Journal of Monetary Economics 56*, 78.

Mariano, R. S. and Y. Murasawa (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics 18*(4), 427–443.

McCracken, M. W. and T. E. Clark (2013). Advances in forecast evaluation. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 2*. Elsevier-North Holland.

Mitchell, J. and K. F. Wallis (2011, 09). Evaluating density forecasts: forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics 26*(6), 1023–1040.

Mitchell, W. C. and A. F. Burns (1938, May). *Statistical Indicators of Cyclical Revivals*. Number mitc38-1 in NBER Books. National Bureau of Economic Research, Inc.

Moench, E. (2012, 06). Term structure surprises: the predictive content of curvature, level, and slope. *Journal of Applied Econometrics 27*(4), 574–602.

Moore, G. H. (1983). The Forty-second Anniversary of the Leading Indicators. In *Business Cycles, Inflation, and Forecasting, 2nd edition*, NBER Chapters, pp. 369–400. National Bureau of Economic Research, Inc.

Quah, D. and T. J. Sargent (1993, June). A Dynamic Index Model for Large Cross Sections. In *Business Cycles, Indicators and Forecasting*, NBER Chapters, pp. 285–310. National Bureau of Economic Research, Inc.

Richard Crump, Stefano Eusepi, D. L. and E. Moench (2014, December). Factor models and mcmc methods for the analysis of the sources and transmission of international shocks. Liberty street economics (blog), Federal Reserve Bank of New York.

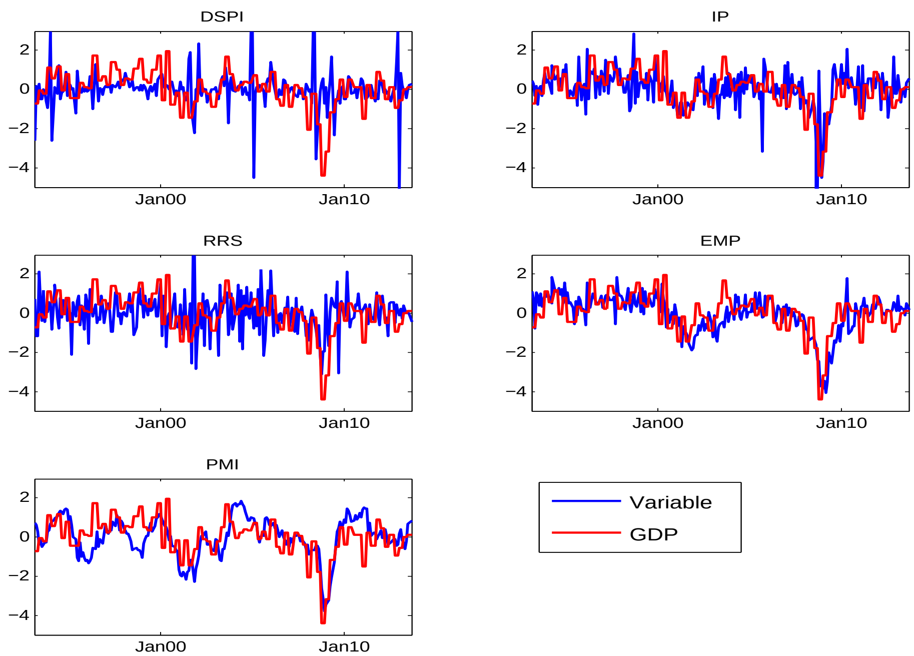
Sargent, T. J. and C. A. Sims (1977). Business cycle modeling without pretending to have too much a-priori economic theory. In C. S. et al. (Ed.), *New Methods in Business Cycle Research*. Federal Reserve Bank of Minneapolis.

Stock, J. H. and M. W. Watson (1992). A probability model of the coincident economic indicators. In G. Moore and K. Lahiri (Eds.), *The Leading Economic Indicators: New Approaches and Forecasting Record*. Cambrigdge University Press.

Stock, J. H. and M. W. Watson (2002, April). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics 20*(2), 147–62.

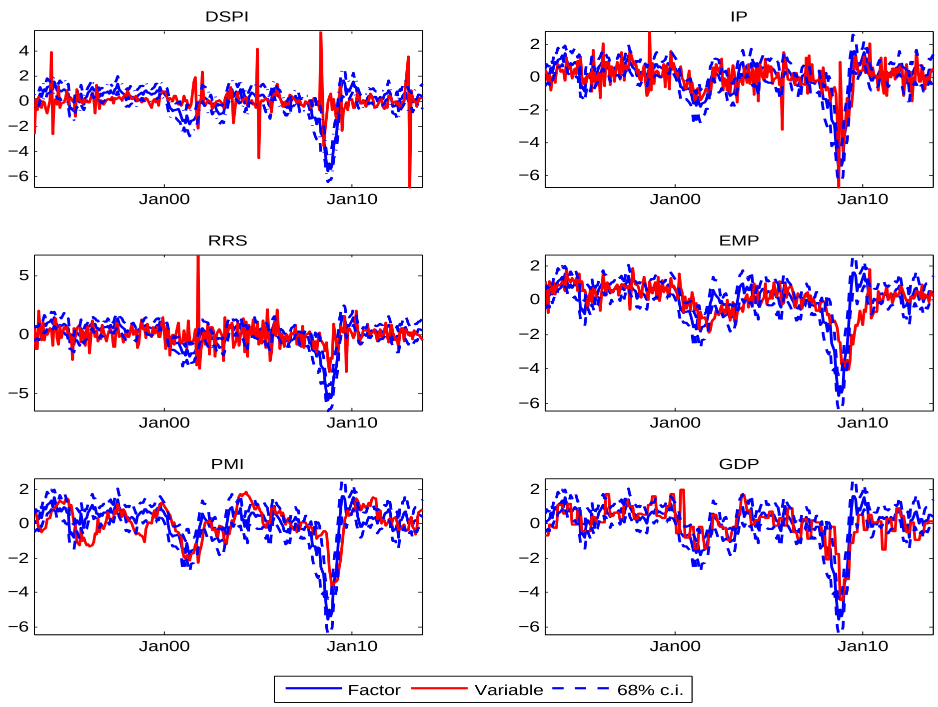
Stock, J. H. and M. W. Watson (2011). Dynamic factor models. In M. P. Clements and D. F. Hendry (Eds.), *Oxford Handbook of Forecasting*. Oxford University Press.

Figure 1: GDP and other variables



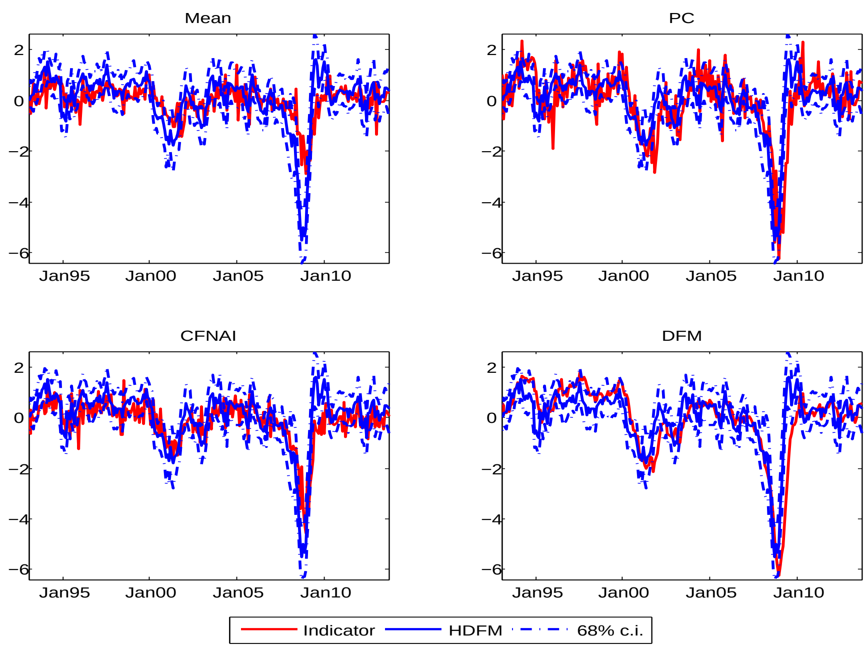
Note: red line: quarter-on-quarter real GDP growth rate; blue line: month-on-month real disposable income growth rate (DSPI), month-on-month industrial production growth rate (IP), month-on-month real retail sales growth rate (RRS), month-on-month employment growth rate (EMP), and level of the purchasing manager index (PMI). Due to the different sampling frequency, GDP growth is reported as constant in the three months of each quarter. Sources: U.S. Bureau of Economic Analysis (BEA), Gross Domestic Product (GDP), Disposable Income Growth Rate (DSPI); Board of Governors of the Federal Reserve System, Industrial Production Index (IP); Federal Reserve Bank of St. Louis, Real Retail Sales Growth Rate (RRS); U.S. Bureau of Labor Statistics, Employment growth rate (EMP); Institute for Supply Management, Purchasing Managers Index (PMI).

Figure 2: Common factor and variables



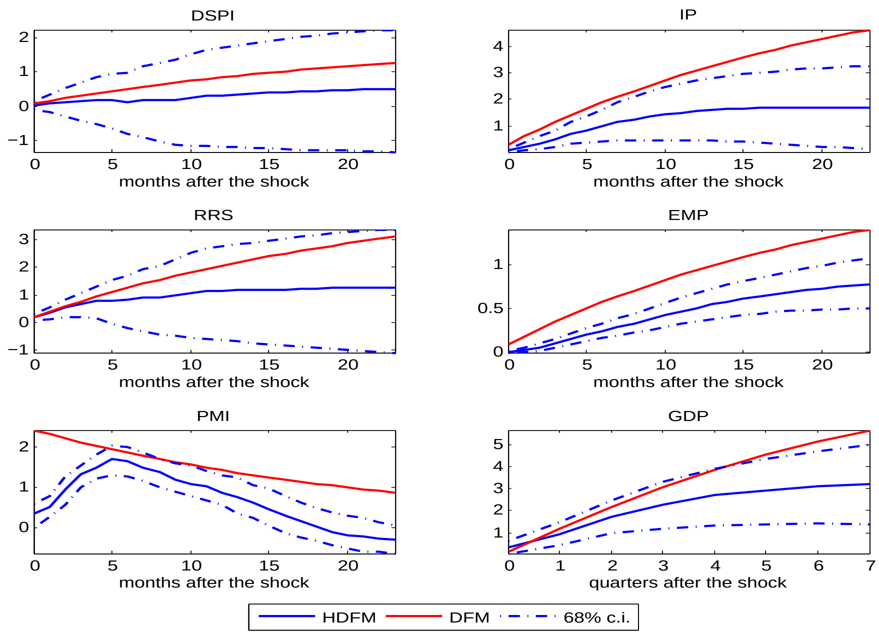
Note: Blue lines: HDFM business cycle indicator (median solid line, 16th and 84th quantiles dashed lines); Red line: month-on-month real disposable income growth rate (DSPI), month-on-month industrial production growth rate (IP), month-on-month real retail sales growth rate (RRS), month-on-month employment growth rate (EMP), the level of the purchasing manager index (PMI), and the quarter-on-quarter real GDP growth rate, reported as a constant in the three months of each quarter. Sources: U.S. Bureau of Economic Analysis (BEA), Gross Domestic Product (GDP), Disposable Income Growth Rate (DSPI); Board of Governors of the Federal Reserve System, Industrial Production Index (IP); Federal Reserve Bank of St. Louis, Real Retail Sales Growth Rate (RRS); U.S. Bureau of Labor Statistics, Employment growth rate (EMP); Institute for Supply Management, Purchasing Managers Index (PMI).

Figure 3: HDFM and other indicators



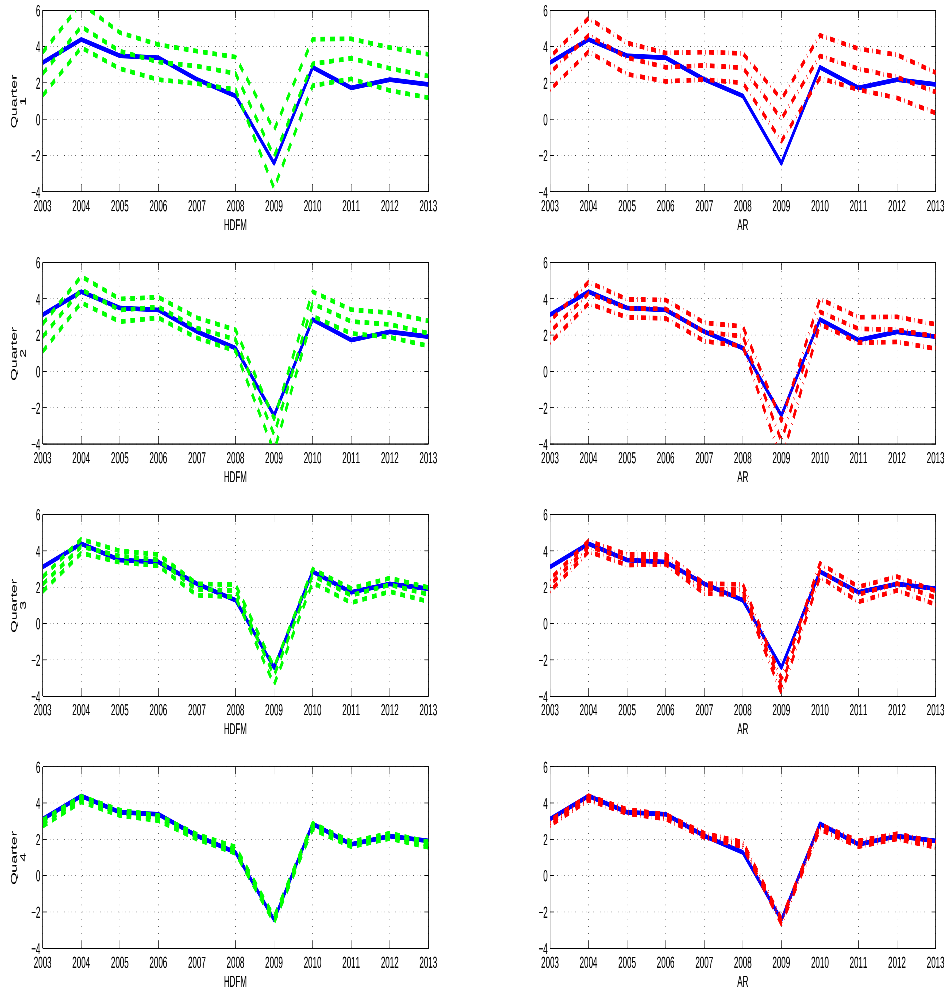
Note: Blue lines: HDFM business cycle indicator (median solid line, 16th and 84th quantiles dashed lines); Red line: simple average of the variables (mean), first principal component of the variables (PC), the Chicago Fed national activity index (CFNAI) and the factor extracted from the homogeneous dynamic factor model (DFM).

Figure 4: IRFs of all variables to a common shock



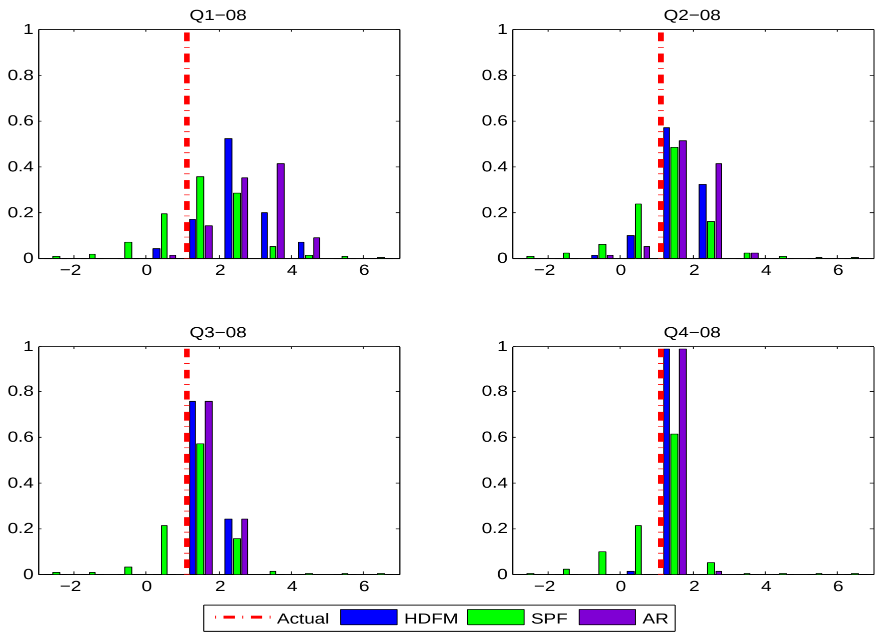
Note: Blue lines: HDFM IRF of the log-levels of the variables (except for PMI, for which we report levels) to a common shock (median solid line, 16th and 84th quantiles dashed lines); Red line: DFM IRF of the log-levels of the variables (except for PMI, for which we report levels) of the variables to a common shock (median).

Figure 5: Nowcasts for the calendar year - 2003 to 2013



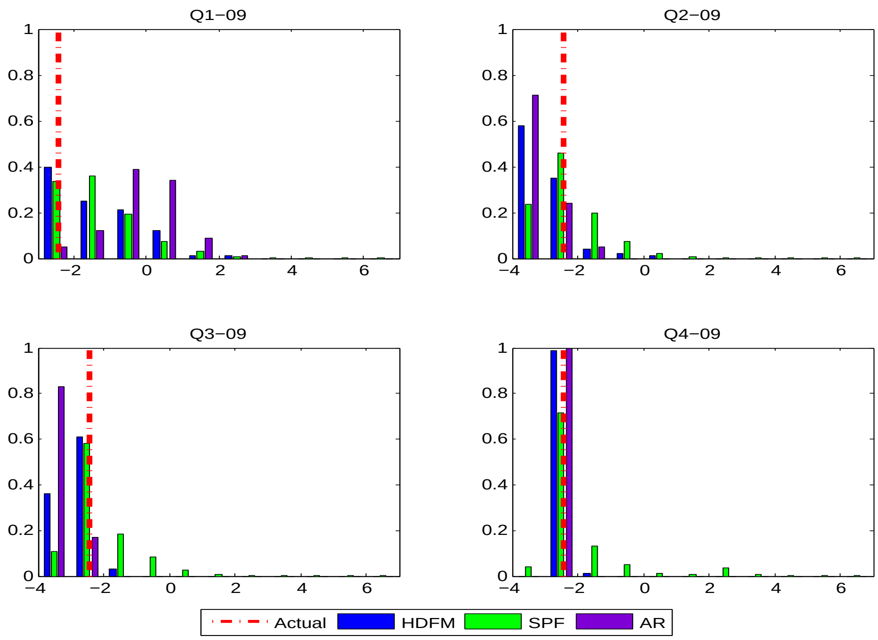
Note: Left panels: HDFM nowcasts (dashed green) and out-turns (solid blue). Right panels: AR nowcasts (dashed red) and out-turns (solid blue). All panels report median, 16th and 84th quantiles of the density nowcasts. From top to bottom, nowcasts produced in quarter 1, 2, 3 and 4.

Figure 6: Case study - calendar year 2008



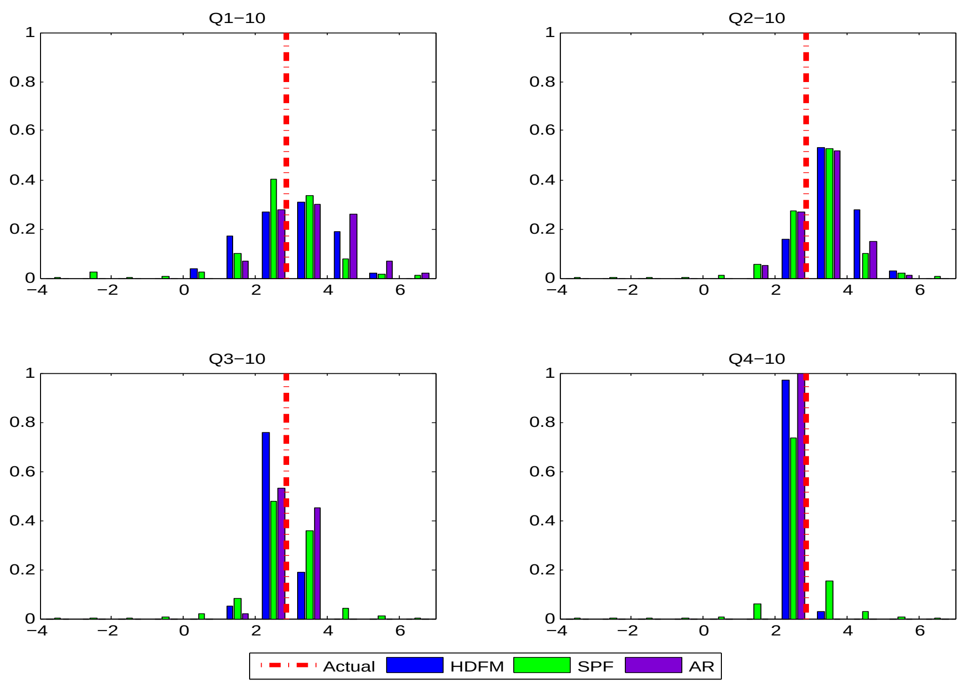
Note: Red dashed line: observed calendar year real GDP growth rate; x-axis: SPF bins; Blue bars: probabilities assigned by the HDFM to the bins; Purple bars: probabilities assigned by the AR to the bins; Green bars: probabilities assigned by the SPF to the bins.

Figure 7: Case study - calendar year 2009



Note: Red dashed line: observed calendar year real GDP growth rate; x-axis: SPF bins; Blue bars: probabilities assigned by the HDFM to the bins; Purple bars: probabilities assigned by the AR to the bins; Green bars: probabilities assigned by the SPF to the bins.

Figure 8: Case study - calendar year 2010



Note: Red dashed line: observed calendar year real GDP growth rate; x-axis: SPF bins; Blue bars: probabilities assigned by the HDFM to the bins; Purple bars: probabilities assigned by the AR to the bins; Green bars: probabilities assigned by the SPF to the bins.