International Journal of Forecasting 35 (2019) 555–572

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Nowcasting and forecasting GDP in emerging markets using global financial and macroeconomic diffusion indexes

Oguzhan Cepni [a], I. Ethem Güney [a,*], Norman R. Swanson [b]

[a] Central Bank of the Republic of Turkey, Anafartalar Mah. Istiklal Cad. No:10 06050 Ulus, Altındag, Ankara, Turkey
[b] Department of Economics, 75 Hamilton Street, New Brunswick, NJ, 08901, USA

ARTICLE INFO

Keywords:
Diffusion index
Dimension reduction methods
Emerging markets
Factor model
Forecasting
Variable selection

ABSTRACT

This paper contributes to the nascent literature on nowcasting and forecasting GDP in emerging market economies using big data methods. This is done by analyzing the usefulness of various dimension-reduction, machine learning and shrinkage methods, including sparse principal component analysis (SPCA), the elastic net, the least absolute shrinkage operator, and least angle regression when constructing predictions using latent global macroeconomic and financial factors (diffusion indexes) in a dynamic factor model (DFM). We also utilize a judgmental dimension-reduction method called the Bloomberg Relevance Index (BRI), which is an index that assigns a measure of importance to each variable in a dataset depending on the variable's usage by market participants. Our empirical analysis shows that, when specified using dimension-reduction methods (particularly BRI and SPCA), DFMs yield superior predictions relative to both benchmark linear econometric models and simple DFMs. Moreover, global financial and macroeconomic (business cycle) diffusion indexes constructed using targeted predictors are found to be important in four of the five emerging market economies that we study (Brazil, Mexico, South Africa, and Turkey). These findings point to the importance of spillover effects across emerging market economies, and underscore the significance of characterizing such linkages parsimoniously when utilizing high-dimensional global datasets.

1. Introduction

Unlike financial variables, which tend to be collected at a higher frequency and published in a timely manner, initial estimates of GDP growth are often released many weeks after the reference quarter.[1] This lack of timely information means that government institutions, such as central banks, are forced to conduct policy activity without a complete knowledge of the current state of the economy. However, central bankers do have timely information on variables that are released at higher frequencies than the GDP, including asset price data and many monthly macroeconomic and financial indicators. This has led to the flourishing of the nascent literature on nowcasting using high dimensional datasets, which involves predicting the current state of the economy before the official figures are released. The key question that this paper attempts to answer is whether it is possible to produce useful early signals of the current state of the economy before the official figures are released. One practical issue that may hamper our efforts in this regard is the fact that many datasets have missing data at the beginning of the sample, particularly in the case of the emerging market economies that we examine. Thus, the usefulness of so-called big data is not a foregone conclusion. For further discussions of nowcasting using big data, see Banbura, Giannone, Modugno, and Reichlin (2013), Bragoli, Metelli, and Modugno (2014), Caruso (2017), Dalhaus, Guénette, and Vasishtha

* Corresponding author.
E-mail addresses: Oguzhan.Cepni@tcmb.gov.tr (O. Cepni), Ethem.Guney@tcmb.gov.tr (I.E. Güney), nswanson@econ.rutgers.edu (N.R. Swanson).

1 The reference quarter is the calendar date to which the data pertain.

https://doi.org/10.1016/j.ijforecast.2018.10.008
0169-2070/

(2017), Foroni and Marcellino (2014), Hindrayanto, Koopman, and de Winter (2016), Luciani, Pundit, Ramayandi, and Veronese (2017), and Modugno, Soybilgen, and Yazgan (2016), as well as the references cited therein.

Given the rich variety of high dimensional datasets that nowcasters now use, it is not surprising that big data methods have come to play an important role in macroeconomic prediction. Indeed, dimension-reduction, shrinkage and machine learning methods are utilized in various of the papers mentioned above. Our objective in this paper is to add to this literature by investigating whether such methods are useful when studying emerging market (EM) economies, including Brazil, Indonesia, Mexico, South Africa, and Turkey. We start by analyzing the usefulness of various dimension-reduction, machine learning and shrinkage methods. These so-called dimension-reduction methods are used for constructing targeted sets of predictors, and include: sparse principal component analysis (SPCA), the elastic net (ENET), the least absolute shrinkage operator (LASSO), and least angle regression (LARS). For further discussions of dimension-reduction methods and prediction, see Bai and Ng (2008), Banbura and Rünstler (2011), Boivin and Ng (2006), Bulligan, Marcellino, and Venditti (2015), Kim and Swanson (2014, 2018), Schumacher (2007), and the references cited therein.[2]

In addition to the statistical procedures above, we also utilize a judgmental variable selection method that yields the so-called Bloomberg Relevance Index (BRI). In particular, Bloomberg reports a "relevance index" that increases monotonically with the number of subscribers to new releases of key economic variables. Put differently, this index assigns a measure of importance to each variable that reflects its usage by market participants. Since the variables that are deemed important by market participants may be good indicators of the state of the economy, we select variables based on this relevance index (for further discussion, see for example Banbura et al., 2013; Bragoli, 2017; Luciani & Ricci, 2014).

It should be noted that we focus on the prediction of GDP growth for two main reasons. First, the initial quarterly GDP growth releases are subject to substantial differences in publication lags for developed and emerging market economies. For example, while the first estimates of GDP in the Euro area and the U.S. are available three and four weeks after the quarter ends, the initial GDPs for Brazil and Turkey are not released until 10 and 12 weeks after the quarter ends, respectively. As the GDP plays a central role in guiding economic decision-making and policy analysis, the construction of timely short-term nowcasts of GDP is quite crucial in the decision-making process of EM central banks. Second, emerging market economies present additional challenges, as the data are often scant and unreliable. When the data quality is low (compared with that available in developed market economies), a careful selection of predictive indicators is crucial in the construction of predictions. Thus, it remains to be seen whether the results found in the literature concerning the usefulness of dimension-reduction methods and diffusion indexes (see e.g. Bulligan et al., 2015; Kim & Swanson, 2014, 2018) carry over to the case of EM countries.

In summary, our approach in this paper involves utilizing the dynamic factor modeling (DFM) framework introduced by Giannone, Reichlin, and Small (2008) in order to construct monthly predictions of the quarterly GDP growth, using a high-dimensional global dataset. More specifically, predictions are constructed using either the entire high dimensional dataset including all variables for a country (or for a group of countries), or only targeted predictors, where targeting is carried out using the SPCA, ENET, LASSO, LARS, and BRI methods discussed above. Thus, our objective is to examine the relevance of two alternative types of diffusion indexes. One variety utilizes only country-specific data, both targeted and un-targeted; while the other utilizes our entire global dataset, both targeted and un-targeted. This approach builds on earlier work by Caruso (2017), Eickmeier and Ng (2011), Foroni and Marcellino (2014), Schumacher (2010), and on a consideration of the usefulness of global high-dimensional datasets for predicting growth in emerging market economies.

Our empirical findings can be summarized as follows. First, when backcasting, nowcasting, and forecasting, there is a substantial reduction in mean square forecast errors (MSFEs) as more data related to the current quarter become available. Thus, the DFM methodology incorporates new information adequately, even for EM countries with data quality issues. This conclusion and our subsequent findings are based on a point forecast evaluation. It is important to note that a density forecast evaluation might yield further insights into the usefulness of the methodology discussed in this paper. Moreover, related empirical research in the area of model confidence sets should prove interesting when examining the robustness of our findings. For a key paper in this area, refer to the study by Hansen, Lunde, and Nason (2011).

Second, predictions based on dimension reduction, machine learning and shrinkage work for EM countries. In particular, benchmark time series models (e.g. autoregressive models) and DFMs that do not utilize dimension reduction yielded inferior predictions in almost all of the prediction experiments that we ran. Interestingly, the BRI method ranks as the best dimension-reduction method, with SPCA coming in second. Together, these targeted predictor selection methods yield MSFE-"best" models around 80% of the time. More specifically, when comparing results across

[2] Bai and Ng (2002, 2006) and Doz, Giannone, and Reichlin (2011, 2012) prove that diffusion indexes are consistent for large values of N and T, where N denotes the number of variables and T is the sample size, in a variety of factor modeling frameworks. Therefore, the common view among practitioners previously was that the dataset with the largest number of indicators should be used for forecasting macroeconomic variables, since it could be argued that leaving out variables might result in a loss of potentially useful information about the state of the economy. Also, the use of many variables reflects a central bank's motivation to prove that it is taking all potentially relevant information into account (see Bernanke & Boivin, 2003). However, a recent branch of the literature has questioned the usefulness of "too much information" in factor model forecasting. For example, Boivin and Ng (2006) show that the use of a small set of appropriately chosen indicators improves macroeconomic forecasts. They select their indicators using the LASSO. Kim and Swanson (2014) and Stock and Watson (2012) discuss the use of other dimension-reduction methods in this context, and also find that forecasts can be improved using targeted predictors.

different prediction horizons and across countries, SPCA yields MSFE-best predictions in 14 of 50 cases (BRI "wins" in 23 out of 50 cases). Thus, while not a crowd-source type of dimension reduction like BRI, the SPCA method clearly performs well, particularly given that it is a purely data-driven statistical learning method. In addition, it is worth noting that our non-BRI dimension-reduction methods, which are all purely statistical, perform their best for nowcasts and backcasts. Indeed, they are MSFE-best in 15 cases out of 20 when considering only nowcasts and backcasts. Thus, the expert judgment associated with using the BRI index is less useful for near-term forecasting, relative to SPCA, ENET, LASSO, and LARS.

Finally, models that include our global EM diffusion indexes are usually MSFE-best for all forecast horizons, as well as across all dimension-reduction methods. For example, in the case of Brazil, global EM diffusion indexes are included in the MSFE-best model for every prediction horizon and across every dimension-reduction method. The picture is similarly clear for Mexico, South Africa, and Turkey: global EM diffusion indexes yield substantial predictive gains, as "Local" and "Local-AR" are the MSFE-best models in only 13 of 30 cases, across all dimension-reduction methods for these countries. Interestingly, the same cannot be said for Indonesia, as "Local" and "Local-AR" are the MSFE-best models in 6 out of 10 cases across all dimension-reduction methods. In summary, we have very strong evidence that global EM diffusion indexes have useful predictive content, suggesting that linkages across EM economies can be modeled using diffusion indexes, and are useful for predicting GDP growth in emerging market economies.

The rest of the paper is structured as follows. Section 2 describes the dataset utilized in our experiments. Section 3 outlines the empirical methodology used in the sequel. This includes discussions of dynamic factor models, dimension-reduction methods, and the setup used in our prediction experiments. Section 4 contains a discussion of our empirical findings. Finally, concluding remarks are collected in Section 5.

2. Data

The dataset used in this paper includes a relatively large set of economic indicators, consisting of 103, 103, 117, 110 and 88 economic series for Turkey, Brazil, Mexico, South Africa, and Indonesia respectively. These series are selected to represent broad categories of economic indicators. Examples of the variables include supply-side indicators, such as industrial production indexes, and demand-side indicators, such as electricity consumption. Various survey variables are also included in the dataset, such as the Markit PMI survey, which is one of the most watched business cycle indicators currently available. Given the sensitivity of emerging market economies to external conditions, we also include value and volume indexes of exports and imports, as well as real effective exchange rates. All data were downloaded from Bloomberg, and a complete list of variables is available at http://econweb.rutgers.edu/nswanson/papers.htm, as well as in Tables B1-B5 in the online appendix.

More specifically, the dataset covers the period January 2005–September 2017 and can be divided into six categories:

1. *Housing and order variables*: house price index, completed buildings recorded and new orders.
2. *Labor market variables*: employment and unemployment.
3. *Prices*: producer prices and consumer prices.
4. *Financial variables*: interest rates, exchange rates, and stock prices.
5. *Money, credit and quantity aggregates*: money supply, mortgage loans, time and sight deposits.
6. *Real economic activity*: PMI survey, industrial production, retail sales and capacity utilization.

All series are made stationary by differencing or log-differencing, as needed. With regard to the timing of data releases, note that survey variables and nominal indicators are usually released during the reference month (i.e., calendar date of the observation), while real and labor variables are released with publication lags of 1–3 months.

Of final note is the fact that the data that we examine are not real-time, in the sense that we do not analyze a sequence of revisions for each calendar-dated observation. Rather, we assume that all data are final revisions. In this sense, our experiments are only pseudo real-time in nature. The construction of real-time datasets for the countries in our analysis, which would enable us to carry out truly real-time prediction experiments, is left to future research.

3. Empirical methodology

3.1. Dynamic factor model

The starting point of our analysis is the widely-used dynamic factor model (DFM) of Giannone et al. (2008). As is typical in such models, individual variables are represented as the sum of components that are common to all variables in the economy (i.e., the factors) and an orthogonal idiosyncratic component. As we shall see later, we also allow for an autoregressive component in our final prediction models.

Formally, the DFM can be written as a system of equations: a measurement equation (i.e., Eq. (1)) that links the observed variables to the unobserved common factor to be estimated, and transition equations (i.e., Eqs. (2) and (3)) that describe the dynamics of the common factor and the residuals of the measurement equation. Once Eqs. (1)–(3) are written in state space form, we utilize the Kalman filter and smoother in order to extract the common factors and generate projections for all of the variables in the model.

In what follows, we consider a panel of observable economic variables $X_{i,t}$, where i indicates the cross-section unit, $i = 1, \dots, N$, and t denotes the monthly time index, $t = 1, \dots, T$. Each variable in the dataset can be decomposed into a common component and an idiosyncratic component, where the common components capture comovements in the data and are driven by a small number of shocks. To summarize, the dynamic factor model can be written as:

$$X_t = \Lambda F_t + \xi_t, \qquad \xi_t \sim N(0, \Sigma_e), \tag{1}$$

Table 1
Selected (r, p) values by predictor selection method and country.

	ALL	BRI	LASSO	ENET	LARS	SPCA
Brazil	1,1	1,1	1,1	1,1	1,1	1,1
Indonesia	2,2	1,2	1,2	1,2	2,1	2,1
Mexico	1,1	1,2	1,1	1,1	1,1	1,1
S. Africa	1,1	2,1	1,2	1,1	1,1	1,1
Turkey	1,1	2,1	1,1	1,1	1,2	2,1

$$F_t = \sum_{i=1}^{p} \Psi_i F_{t-i} + u_t, \qquad u_t \sim N(0, Q), \tag{2}$$

$$\xi_t = \rho \xi_{t-1} + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma^2), \tag{3}$$

where F_t is an $r \times 1$ vector of unobserved common factors with zero mean and unit variance that reflect most of the co-movement in the variables, Λ is a corresponding $N \times r$ factor loading matrix, and the idiosyncratic disturbances ξ_t are uncorrelated with F_t at all leads and lags, and have a diagonal covariance matrix Σ_e. The number of lags of the common factor is defined as p. It is assumed that the common factors F_t follow a stationary VAR(p) process that is driven by the common shocks, $u_t \sim N(0, Q)$, and that the Ψ_i are $r \times r$ matrices of autoregressive coefficients. Also, the common shocks u_t and the idiosyncratic shocks ϵ_t are assumed to be serially independent and independent of each other over time, while weak cross-correlation and serial correlation in the idiosyncratic shocks are allowed.

We construct forecasts of our quarterly target series, say y_t, in the monthly DFM framework by expressing each quarterly variable in terms of a partially observed monthly series, following the approach of Mariano and Murasawa (2003). That is, we assume that:

$$y_t = \mu + \beta' F_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2). \tag{4}$$

When implementing this model, we use recursively-estimated parameters that are updated on a monthly basis, prior to the construction of each new prediction. We select the number of factors (r) and the lag order (p) in Eq. (2) by searching across all combinations of $r = 1, \ldots, 4$ and $p = 1, \ldots, 4$. In particular, we select the model for each economy by comparing the out-of-sample performances for all combinations of parameters.[3] We find that simple model specifications, with one or two factors and one lag, often yield the best out-of-sample performances. In particular, recursive estimation of r and p is not really needed in our experiments, because the "optimal" models have the parameter selections for r and p that are shown in Table 1.[4]

In Table 1, ALL denotes the cases where all variables in our dataset are used in factor construction. In all other cases, dimension-reduction methods are used to construct the sets of predictors used in our experiments. These methods include BRI, LASSO, ENET, LARS and SPCA, all of which are discussed in the next subsection.[5] As is the case in many forecasting studies that have used factor models (see e.g. Kim & Swanson, 2014, and the references cited therein), we find that simple model specifications, with one or two factors and one or two lags, often yield the best nowcasts and forecasts.[6] For a further discussion of the trade-off between using parsimonious one- or two-factor models and models with many factors, in which case the more heavily parameterized models usually lead to poor forecasting performances, see Bragoli (2017), Forni, Hallin, Lippi, and Reichlin (2000), and Stock and Watson (2002).

3.2. Dimension reduction methods for selecting targeted predictors

Since our dataset is characterized by a large number of variables (see Section 2), it is important to select appropriate "targeted" predictors prior to estimating factor models. This is because model and parameter uncertainty can have an adverse impact on the marginal predictive content of factors that are constructed using finite samples of data. Moreover, using least squares or other standard estimators directly is not feasible in the contexts considered in this paper, since the number of regressors that we consider is greater than the number of observations. These issues are discussed at length by Bai and Ng (2008), Kim and Swanson (2014, 2018), Kuzin, Marcellino, and Schumacher (2011, 2013), Stock and Watson (2012), and many others. Thus, we utilize variable selection or dimension-reduction methods in order to pre-select predictors prior to the construction of factors. Many of the shrinkage and machine learning methods that are examined by the above authors in this context can be interpreted as penalized estimation problems. For example, Kim and Swanson (2018) implement a number of interesting variable selection and shrinkage methods, including bagging, boosting, least angle regression, and the non-negative garrote, and find strong evidence of the usefulness of dimension-reduction techniques in the context of out-of-sample forecasting of 11 U.S. macroeconomic variables. Bai and Ng (2008) implement the least absolute shrinkage selection operator and the elastic net in order to construct targeted predictor datasets, and find improvements at all forecast horizons when estimating factors using fewer informative predictors. Also, Bulligan et al. (2015) show that soft thresholding methods can be used successfully to reduce the size of large panels of economic data.

This paper implements a variety of variable selection and shrinkage methods in order to obtain targeted predictors for use in our factor model, including:

- least absolute shrinkage selection operator (LASSO)
- elastic net estimator (ENET)

[3] We also considered an alternative approach where we chose r and p based on the approach of Bai and Ng (2002) and the BIC, respectively. The results based on this approach were subjectively the same as those based on the approach reported here.

[4] Our findings based on setting $(r, p) = (1, 1)$ for all experiments are qualitatively the same as those reported in the following parts.

[5] We also used the Bai and Ng (2002) criterion for selecting r, and found that it chooses more factors than our approach, resulting in a deterioration in forecast accuracy.

[6] A separate recursive principal component analysis that was carried out in order to further investigate the choice of r in our setup found that approximately 75% or more of the variation in GDP growth is explained by the first two common factors for all five emerging market economies analyzed in our experiments.

- least angle regressions (LARS) and
- sparse principal component analysis (SPCA).

Recently, there has been tremendous progress made in the development of interesting new shrinkage methods. For example, various so-called adaptive methods have been introduced and used for the forecasting of macroeconomic variables. For instance, the adaptive LASSO, adaptive elastic net, group LASSO and Bayesian LASSO methods have all received particular attention. Zou (2006) introduced the adaptive LASSO in order to show that the LASSO does not have the oracle property and that its performance tends to deteriorate when the number of variables increases more quickly than the number of observations. Garcia, Medeiros, and Vasconcelos (2017) use high-dimensional machine learning to forecast Brazilian CPI inflation, and find that the LASSO and adaptive LASSO models perform well at shorter horizons. However, both LASSO regressions and elastic net regressions select variables individually, making the interpretation of the final model more difficult. Yuan and Lin (2006) therefore introduce a Group LASSO algorithm in order to impose sparsity constraints at a 'block-variable' level, making it easier to interpret the final models. In related research, Li and Chen (2014) utilize LASSO-based methods for forecasting macroeconomic variables in the U.S., and find that the group LASSO that shrinks variables at the 'block-variable' level results in predictive gains. A comparison of our empirical findings with those based on the use of the methods above should be of great interest to forecasters, although this topic is beyond the scope of the current paper.

Turning again to our empirical setup, we consider a panel of observable economic variables $X_{i,t}$, where i indicates the cross-section unit, $i = 1, \ldots, N$, and t denotes the time index, $i = 1, \ldots, T$, as discussed above. Following the notation of Hastie, Tibshirani, and Friedman (2009), we consider the problem of selecting a subset of X, where X is a $T \times N$ matrix to be used for forecasting the scalar annualized GDP growth, say Y, for $i = 1, \ldots, T$.

3.2.1. Sparse principal component analysis

Sparse principal component analysis (SPCA), introduced by Zou, Hastie, and Tibshirani (2006), is a variant of principal component analysis (PCA). PCA yields orthogonal latent factors that are maximally correlated with all variables in X. One potential disadvantage of this method is that each principal component is a linear weighted combination of all variables in the original dataset, with no weights being equal to zero. Thus, all variables are included in all factors. SPCA, which can be interpreted as double-shrinkage using the elastic net, combines L_1 and L_2 penalty functions (in a penalized regression problem) in order to "shrink" the weights from PCA factors to zero. Thus, the factors constructed using SPCA contain non-zero weights only on selected (or targeted) predictors. Setting various factor loading coefficients equal to zero in this way has the potential to reduce the "noisiness" of the factors, as well as to assist in the economic interpretation of factors.

The SPCA problem can be formulated as the following maximization problem:

$$\begin{aligned} \underset{X}{\text{maximize}} \quad & v^T(X^TX)v, \\ \text{subject to} \quad & \sum_{j=1}^{N} |v_j| \leqslant \psi, \\ & v^Tv = 1, \end{aligned}$$

where X is the data matrix, v are principal components (possibly with zero loadings), and ψ is a tuning parameter. Optimization in this context is not trivial, and the literature has suggested various algorithms based on convex semidefinite programming, generalized power methods, greedy search methods, and exact methods using branch-bound techniques. Following Naikal, Yang, and Sastry (2011), we implement the augmented Lagrange multiplier method for extracting the sparse principal components. In particular, we select the first factor (i.e., the maximal correlation factor), and our targeted predictors are the variables with non-zero factor loading coefficients in said factor.

3.2.2. Least absolute shrinkage operator (LASSO)

We also implement the LASSO, which was introduced by Tibshirani (1996) and can be written as a penalized regression problem, just like the well-known ridge estimator, for example. However, LASSO imposes an ℓ_1-norm penalty on the regression coefficients, rather than an ℓ_2-norm penalty as is the case with the well known ridge estimator. This penalty results in a (possible) shrinkage of the coefficients (called $\hat{\beta}^{lasso}$ below) to zero. The LASSO estimator is

$$\hat{\beta}^{lasso} = \min_{\beta} \quad \|Y - X\beta\|_2 + \lambda \sum_{j=1}^{N} |\beta_j|, \tag{5}$$

where λ is a tuning parameter that controls the strength of the ℓ_1-norm penalty. Since the objective function in the LASSO is not differentiable, numerical optimization must be used when constructing $\hat{\beta}^{lasso}$. For example, an efficient iterative algorithm called the "shooting algorithm" is proposed by Fu (1998). One of the limitations of the LASSO approach is that the number of variables selected is bounded by the sample size. For example, if $N > T$, the LASSO yields at most N non-zero coefficients (see Swanson, 2016, for further discussion). The variables associated with these non-zero coefficients constitute our set of targeted predictors when using the LASSO. Our experiments utilize the algorithm of Fu (1998).

3.2.3. Elastic net (ENET)

The LASSO is naturally adapted to cases where there are many zero coefficients in the "true" model. However, in the presence of highly correlated predictors, Tibshirani (1996) shows that the predictive performance of the LASSO is sometimes worse than that of the forecasts that are constructed using ridge regression. Zou and Hastie (2005) address this issue by proposing a hybrid form of the LASSO and ridge estimators, called the elastic net (ENET) estimator. The ENET estimator is defined as

$$\hat{\beta}^{EN} = \min_{\beta} \quad \|Y - X\beta\|_2 + \lambda_1 \sum_{j=1}^{N} |\beta_j| + \lambda_2 \sum_{j=1}^{N} {\beta_j}^2, \qquad (6)$$

where there are now two tuning parameters controlling the two penalty functions, λ_1 and λ_2. The EN estimator also results in a possible shrinkage of coefficients to zero, although the EN can yield more than N non-zero coefficients in cases where $N > T$.

3.2.4. Least angle regressions (LARS)

Least angle regression was proposed by Efron, Hastie, Johnstone, and Tibshirani (2004). The algorithm is similar to forward step-wise regression, but instead of including variables at each step, the algorithm proceeds equiangularly in directions that are chosen to impose equal correlations with each of the variables currently in the model. Moreover, LARS can be reformulated easily so as to obtain solutions for other estimators, like the LASSO and EN. It allows for the ranking of different predictors according to their predictive content, which is not the case when using hard thresholding methods. Thus, sparsity can be obtained by selecting only the highest ranked variables for model estimation. This paper follows the approach of Efron et al. (2004) when implementing LARS.

3.2.5. Bloomberg relevance index (BRI) for selecting targeted predictors

In addition to the techniques above, we also investigate selected targeted predictors based on observing which economic variables are monitored by the markets. This type of expert judgement method was developed by citetBanburaetal2013 and has been used by Bragoli et al. (2014), Luciani and Ricci (2014), and Luciani et al. (2017). The main assumption of this approach is that market participants monitor macroeconomic data and use them to form their expectations about the state of the economy when allocating their investments. In this context, the Bloomberg reports a "relevance index", which we call the BRI index, for numerous economic variables that are followed closely by market participants. We select the variables based on this index. As Bloomberg only reports current values for this index, all of our BRI targeted predictors are based on the Bloomberg information that was available at the time when our dataset was pulled (i.e., September 2017). We maintain comparability across all of the different predictor selection methods in our experiments by applying all other methods using the same dataset as that available to the users of Bloomberg in January 2018. More specifically, when Bloomberg users put an "alert" on the release date of a variable in the Bloomberg database, the relevance index for that variable increases. For the sake of simplicity, we select all variables that have BRI values bigger than zero. In particular, the BRI index selects 20, 16, 22, 23, and 28 variables for Turkey, Indonesia, Mexico, South Africa and Brazil, respectively.[7]

[7] The targeted predictors used in our experiments are selected only once, based on an analysis of our entire dataset. This is an approximation of the approach that a central bank might take, for example, of selecting a new set of predictors prior to the construction of each prediction, and is predicated on the lack of data availability in our sample period.

Table 2
GDP growth rate correlation coefficients.

	Turkey	Brazil	Indonesia	S. Africa	Mexico
Brazil	0.39	1	0.46	0.67	0.35
Indonesia	0.04	0.46	1	0.24	0.16
Mexico	0.70	0.35	0.16	0.54	1
S. Africa	0.44	0.67	0.24	1	0.54
Turkey	1	0.39	0.04	0.44	0.70

3.2.6. Selected predictors for five emerging market economies

We provide insight into the predictors that are selected using the five above methods by listing the key variables (i.e., those that are chosen for at least four of the five methods above) in Table 3.[8] Several interesting conclusions can be drawn in terms of cross-country differences and similarities when comparing the variables for each of our five countries.

For Turkey, note that many variables are related to industrial production and its subcomponents, all of which play important leading roles in driving cyclical fluctuations in GDP growth. Also, the Turkish economy is driven to a large extent by domestic demand (i.e., consumption expenditures), and imports tend to increase markedly when the economy is in an expansion phase. Thus, it is not surprising that imports, which are good predictors of consumption expenditure, are in the set of selected predictors. Finally, it is worth noting that confidence indexes are important predictors for Turkey, suggesting that these indexes are accurate measures of consumer and producer sentiment.

For Mexico, the selected predictors are mainly export measures. This is not altogether surprising, given that Mexico relies heavily on trade, and is the USA's most important trade partner. Indeed, non-petroleum exports to the US comprise nearly 83% of their total non-petroleum exports. Total vehicle production is another important indicator. The automobile sector in Mexico differs from those of other emerging market countries because it produces technologically complex components, while other countries function as "assembly" manufacturers. Finally, variables related to labor force statistics are important for Mexico too.

Brazil's economy also relies heavily on exports, so it is not surprising that the predictors selected include various trade-related variables. Also, although commodities-related sectors play an important role in Brazil, manufacturing sectors also play a significant role in the economy. Hence, the labor force and working hours variables related to the manufacturing industry are also relevant for GDP growth. Furthermore, three retail sales indexes are key predictors, which is not surprising, since the main driver of the GDP growth is private consumption. Interestingly, two of these retail indexes are construction-related, suggesting that government efforts toward revitalization through the implementation of urbanization programs are an important driver of growth in Brazil.

A similar pattern emerges when looking at the variables selected for Indonesia; that is, exports and imports matter, as do a number of retail sales indexes. On the other hand,

[8] The full list of selected predictors, by method, is available from the authors upon request.

Table 3
Key predictors selected using dimension-reduction methods.

Mexico	Brazil
Exports by Sector Non-Petroleum	Manufacture Industry Employment
Trade Balance Exports	Manufacture Industry Working Hours
Vehicle Production	Industry Confidence General
Vehicle Sales	Trade Balance FOB Imports
Vehicle Exports Total	Export Price Index
Non-Manufacturing Index New Orders	Retail Sales Volume
Manufacturing Index New Orders	Retail Sales Volume Construction Materials
Unemployment Rate	Retail Sales Volume Furniture& Domestic Appliance
Employment Rate	Family Consumption
Formal Job Temporary&Permanent Workers Manufacturing	
Turkey	**S.Africa**
Industrial Production	Manufacturing SA Constant Prices
Industrial Production: Intermediate Goods	Wholesale Retail Hotels SA Constant Prices
Industrial Production: Capital Goods	Electricity SA Constant Prices
Industrial Production: Manufacturing	Real GDP Expenditure on GDP
Capacity Utilization	Composite Business Cycle Indicator - Coincident Indicator
Real Sector Confidence Index	FTSE/JSE Africa Industrials Index
Real Sector Confidence Index: Volume of Orders (Current Situation)	FTSE/JSE Africa All Share Index
Real Sector Confidence Index: Export Orders (Next 3 Months)	FTSE/JSE Africa Top40 Tradeable Index
GDP Transportation & Storage Constant Prices	FTSE/JSE Africa Basic Materials Index
GDP Final Consumption Expenditure of Residents	Bloomberg South Africa Exchange Market Capitalization USD
Turkey Trade Imports WDA	
Non Agricultural Unemployment Rate	
Indonesia	
GDP Current Prices Expenditure Exports Goods & Services	
GDP Current Prices Expenditure Import Goods & Services	
Exports	
Exports: Oil & Gas	
Imports: Oil & Gas	
Gaikindo Motor Vehicle Local Car Sales	
Asosiasi Industri Sepedamotor Local Number of Motorcycles Sold	
Wage for Construction Worker per Day Nominal	
Wage for Household Servant per Month Nominal	
External Debt Total	
Working Capital Loans Total	
Non-Performing Loan (Gross)	

the predictors for Indonesia include not only real variables, but also financial variables, such as external debt and loans. This is not surprising for an emerging market economy, since capital loan growth is a key indicator of new investment, which in turn is a predictor of GDP growth.

For South Africa, we also observe that financial variables are important. One reason for this may be that levels of domestic savings are inadequate, resulting in a heavy reliance on capital in-flows in order to spur economic growth.

3.3. Prediction experiments

We evaluate the forecasting performance of the above dynamic model by using a recursive forecasting scheme, expanding the model estimation sample prior to the construction of each new forecast. The estimation sample starts in January 2005, and our out-of-sample evaluation period is July 2008–September 2017. Put differently, we carry out a series of recursive pseudo out-of-sample forecasting experiments for the prediction period July 2008–September 2017, where monthly forecasts for the five emerging market economy GDP growth rates are constructed. In addition, the experiments are repeated using the various different dimension-reduction methods discussed above. For each reference quarter (recall that the GDP is measured quarterly), we produce a sequence of ten monthly predictions, starting with a forecast based on information available in the first month of the two previous quarters and ending with a forecast based on information available in the first of month of the subsequent quarter before the GDP is actually released. Thus, we construct three monthly two-quarter-ahead forecasts ($h = 2$), three monthly one-quarter-ahead forecasts ($h = 1$), three monthly nowcasts ($h = 0$), and one monthly backcast ($h = -1$) for a quarterly forecast of GDP.

We carry out two varieties of experiment. In our first set of experiments, we investigate the forecasting performances of DFM predictions directly based on a pre-selection of predictors from large panels of macroeconomic data (see Section 3 for a discussion of the data used). That is, the following five dimension-reduction methods are utilized for each country in order to select predictors for inclusion in the DFM model: BRI, LASSO, ENET, LARS, and SPCA. Two benchmark models are also used to construct predictions, including an autoregressive (AR) model with lags selected via the Schwarz information criterion (SIC), and a version of our DFM model, called "ALL" in Table 4, where factors are extracted using all of the domestic variables for a given country. A comparison of our targeted predictor results with AR and ALL allows us to assess the predictive accuracy relative to that of a standard strawman model that is used widely in the literature (i.e., the AR model), as well as with a factor model where predictors are not targeted (i.e., the ALL model).

Our second set of experiments involves combining the targeted predictors used in our first set of experiments across all five countries. This resulting set of "Global" targeted predictors is then partitioned into three sets of variables: "Global" (includes all variables), "Macroeconomic" (includes only macroeconomic variables) and "Financial" (includes only financial variables). These subsets of variables are used individually to specify new factors that are

Table 4
Mean square forecast errors (MSFEs) based on the use of different dimension-reduction and shrinkage methods.

Brazil	Forecast ($h = 2$)			Forecast ($h = 1$)			Nowcast ($h = 0$)			Backcast ($h = -1$)
	1	2	3	1	2	3	1	2	3	1
AR	4.48	4.48	4.26	3.73	3.73	3.39	2.84	2.84	2.45	2.45
ALL	1.01	0.92	0.89	0.83	0.73*	0.67**	0.52	0.38*	0.29*	0.42*
BRI	0.94	**0.84**$_{LB}$	**0.83**$_{LB}$	0.81	**0.66**$_{LB}^{**}$	**0.63**$_{LB}^{*}$	0.53*	0.36*	0.34*	0.55*
LASSO	**0.94**$_{LB}$	0.87	0.87	**0.79**$_{LB}^{*}$	0.69*	0.67**	**0.50**$_{LB}^{*}$	0.33*	0.24*	**0.33**$_{LB}^{*}$
ENET	0.97	0.90	0.90	0.83*	0.71*	0.69**	0.54*	0.35*	0.27*	0.35*
LARS	0.97	0.90	0.85	0.81	0.70*	0.63**	0.50*	0.35*	0.24*	0.44*
SPCA	0.97	0.89	0.88	0.81	0.69*	0.65*	0.50*	**0.32**$_{LB}^{*}$	**0.23**$_{LB}^{*}$	0.36*
Indonesia										
AR	1.11	1.11	1.05	1.09	1.09	1.01	0.83	0.83	0.75	0.75
ALL	1.23	1.47	1.16	1.21	1.36	0.97	1.20	1.12	0.73*	0.84
BRI	**0.76**$_{LB}$	**0.75**$_{LB}$	**0.78**$_{LB}$	**0.72**$_{LB}^{*}$	**0.70**$_{LB}^{*}$	**0.69**$_{LB}^{*}$	0.82	0.81	0.81	0.89
LASSO	0.85	0.84	0.87	0.80	0.80	0.75	0.94	0.96	0.88	0.94
ENET	0.98	1.05	0.95	0.86	0.89	0.78*	1.00	1.05	0.87	0.95
LARS	1.10	1.12	1.15	0.98	1.02	1.02	0.98	0.95	1.09	1.18
SPCA	1.06	0.99	1.21	0.80	0.74	0.77	**0.75**$_{LB}$	**0.73**$_{LB}$	**0.70**$_{LB}$	**0.78**$_{LB}$
Mexico										
AR	4.48	4.48	3.99	3.65	3.65	3.07	2.59	2.59	1.97	1.97
ALL	0.90	0.78	0.91	0.70*	0.58*	0.61*	0.49**	0.39**	0.37*	0.42*
BRI	**0.56**$_{LB}$	**0.54**$_{LB}$	**0.66**$_{LB}$	**0.64**$_{LB}^{**}$	**0.54**$_{LB}^{*}$	**0.52**$_{LB}^{**}$	**0.45**$_{LB}^{**}$	0.38**	**0.26**$_{LB}^{*}$	0.49*
LASSO	0.88	0.77	0.87	0.71*	0.60*	0.61*	0.52**	0.41**	0.34*	**0.38**$_{LB}^{*}$
ENET	0.85	0.75	0.84	0.69*	0.59*	0.59*	0.52**	0.43**	0.36*	0.42*
LARS	0.83	0.72	0.81	0.66*	0.55*	0.55*	0.48**	0.39**	0.31*	0.39*
SPCA	1.09	0.92	1.17	0.74	0.58*	0.73	0.45**	**0.33**$_{LB}^{**}$	0.51*	0.56*
South Africa										
AR	2.68	2.68	2.47	2.20	2.20	1.96	1.58	1.58	1.32	1.32
ALL	0.91	0.85	0.83	0.81*	0.74**	0.66*	0.71**	0.64**	0.46**	**0.43**$_{LB}^{**}$
BRI	0.82	0.75	0.83	0.66**	0.58**	0.65*	0.43**	**0.37**$_{LB}^{**}$	**0.41**$_{LB}^{**}$	0.49**
LASSO	0.82	0.82	0.70	0.74*	0.74	0.57*	0.64**	0.65*	0.47**	0.55**
ENET	0.81	0.76	0.73	0.70**	0.64**	0.55*	0.63**	0.59**	0.46**	0.52**
LARS	0.86	0.80	0.78	0.75*	0.69**	0.61*	0.70**	0.67**	0.60**	0.68**
SPCA	**0.58**$_{LB}$	**0.51**$_{LB}$	**0.49**$_{LB}$	**0.43**$_{LB}^{**}$	**0.43**$_{LB}^{**}$	**0.42**$_{LB}^{*}$	**0.43**$_{LB}^{**}$	0.45**	0.51**	0.56**
Turkey										
AR	7.58	7.58	7.00	6.48	6.48	5.77	4.95	4.95	4.17	4.17
ALL	0.82	0.75	0.80	0.73**	0.65*	**0.62**$_{LB}^{*}$	0.63	0.60*	0.59*	0.70
BRI	0.82	0.72	**0.74**$_{LB}$	**0.53**$_{LB}^{*}$	**0.53**$_{LB}^{*}$	0.72*	0.60	0.61*	0.65*	0.66*
LASSO	0.90	0.82	0.88	0.77	0.68*	0.70	0.57	0.50*	0.51*	0.68
ENET	**0.74**$_{LB}$	**0.70**$_{LB}$	0.76	0.70**	0.64**	0.65*	0.68	0.67	0.70	0.83
LARS	1.24	0.88	0.87	0.87	0.65	0.64	**0.46**$_{LB}^{*}$	**0.41**$_{LB}^{*}$	**0.48**$_{LB}^{*}$	**0.53**$_{LB}^{*}$
SPCA	0.85	0.72	0.86	0.70*	0.60*	0.68	0.54*	0.50*	0.66	0.79

Notes: The entries are MSFEs, with the method that yields the smallest MSFE being shown in bold. The entries in the first row correspond to actual point MSFEs of our benchmark AR(SIC) model, while the remaining entries are relative MSFEs (i.e., relative to the AR(SIC) benchmark model). Thus, a value below unity indicates that the dynamic factor model point MSFE for a particular dimension-reduction method (listed in the first column) is more accurate than that based on the AR(SIC) benchmark. For each country, entries that are highlighted and have a "LB" subscript added are the MSFE-best models across all dimension-reduction methods for a given quarterly forecast horizon (i.e., $h = -1, 0, 1$, or 2) and forecast month within the quarter (i.e., 1, 2, or 3). Entries marked with an asterisk(s) (** 5% level; * 10% level) are significantly superior to the AR(SIC) benchmark model, based on the application of the DM predictive accuracy test discussed in Section 3.3. See Section 3 for complete details.

"added" to our DFM model. In particular, three new diffusion indexes (i.e., factors) are constructed for each economy. These new diffusion indexes are called the EM Global factor (constructed using "Global" variables), the EM Macro factor (constructed using "Macroeconomic" variables), and the EM Financial factor (constructed using "Financial" variables).[9] The new factors are then included in the following four specifications, in which Specification 1 is simply the model in Eq. (4), and Specifications 2–4 are extensions that include our new diffusion indexes.

- **Specification 1**: Local diffusion index model $y_{t+h} = \mu + \beta' F_t^{Local} + \varepsilon_{t+h}$
- **Specification 2**: EM Global factor model $y_{t+h} = \mu + \beta' F_t^{Local} + \vartheta' F_t^{EMGlobal} + \varepsilon_{t+h}$
- **Specification 3**: EM Macro factor model $y_{t+h} = \mu + \beta' F_t^{Local} + \theta' F_t^{EMMacro} + \varepsilon_{t+h}$
- **Specification 4**: EM Macro-Financial factor model $y_{t+h} = \mu + \beta' F_t^{Local} + \theta' F_t^{EMMacro} + \delta' F_t^{EMFinancial} + \varepsilon_{t+h}$

9 Note that the cross-country diffusion indexes (i.e., factors) that we extract for each country do not include local variables from the corresponding country for which the new factors are being constructed.

Finally, we analyse four additional variants of Specifications 1–4 that include lags of y_t, with lags selected via the SIC. Moreover, as was done in our first set of experiments, we also construct predictions using a purely autoregressive model, referred to above as AR.

We assess the precision of the different sequences of forecasts constructed in the experiments above using the mean square forecast error (MSFE), which is measured as the average of the squared differences between the predicted and actual GDP growth rates. We also assess the statistical significance of differences in MSFEs across models and methods by conducting predictive accuracy tests using the Diebold and Mariano (1995, DM) test, which is implemented using quadratic loss, and which has a null hypothesis that the two models being compared have equal predictive accuracies. For a complete discussion of inference based on the DM test both in cases where models are nested and in cases where parameter estimation error is accounted for in the limit distribution of the test statistic, refer to the studies by McCracken (2000) and Corradi and Swanson (2006, 2007).

4. Empirical results

4.1. Mining big data using dimension-reduction methods

Table 4 summarizes the results of our first set of prediction experiments, in which we compare the five dimension-reduction methods that are used for selecting targeted predictors prior to the construction of predictions using dynamic factor models. The table is partitioned vertically into five sets of results for our five EM economies. The entries in the table are either MSFEs (for the AR(SIC) model listed in the first row for each country) or relative MSFEs (for all other rows under each country), where the numerator in the relative MSFEs is the MSFE of the benchmark AR(SIC) model. In particular, the entries are MSFEs for various types of predictions. For each of two quarterly h-step-ahead forecast horizons (i.e., $h = 1$ and $h = 2$), MSFEs from three monthly forecasts (denoted as months "1", "2" and "3" of the quarter in the second row of the table). Results are also reported for three monthly nowcasts (for the quarterly forecast horizon $h = 0$), and for one monthly backcast (for the quarterly forecast horizon $h = -1$). For each country, entries shown in bold and with a subscript "LB" (for "locally-best") indicate the MSFE-"best" models across all dimension-reduction methods, for a given forecast horizon and country. A summary of the"best" dimension-reduction methods from this table is given in Table A.1 of online appendix.

The results in Table 4 reveal various interesting insights.

First, there is a substantial reduction in MSFEs as more data related to the current quarter become available, as is clearly evident when one scans the rows of the table from left to right as one moves from forecasting (least information) to backcasting (most information). Thus, the DFM is incorporating new information effectively (see Banbura & Rünstler, 2011; and Giannone et al., 2008; for further discussion).

Second, with a limited number of exceptions, the entries in Table 4 are all less than unity, which indicates that our predictions are quite accurate relative to the benchmark model. In addition, the magnitudes of the MSFEs are similar for most countries, except for Turkey, where the errors are much larger due to a higher GDP growth volatility.

Third, recall that there are ten forecast horizons and five countries, meaning that we have a total of 50 specifications for each dimension-reduction method. Of the various methods, the BRI criterion performs surprisingly well, as it attains the top rank in 23 cases out of 50. This can be seen from Table 4 by noting that the lowest MSFEs are shown in bold. Indeed, the average MSFEs of BRI-type predictions are 31%, 23%, 47%, 36%, and 34% lower than those associated with the AR model for Brazil, Indonesia, Mexico, South Africa, and Turkey, respectively. Of note is that Bloomberg collects forecasts from market analysts in order to produce their own GDP growth forecasts (generally around two weeks before the release of new GDP data). Their predictions are revised continually up to 24 h before the release of actual data. This implies that market analysts monitor all macroeconomic data continually in order to form expectations on current and future GDP growth values, and this monitoring behavior is reflected in the BRI index that we use, since it is based on user subscriptions to Bloomberg news alerts for specific data releases of variables that are deemed important. Thus, the BRI index can be interpreted as a form of big data based crowd-sourcing information.

Fourth, SPCA also fares quite well, yielding MSFE-best predictions in 14 of 50 cases. While not a crowd-source type of dimension reduction like the BRI index, the SPCA method clearly performs well, particularly given that it is a purely data-driven statistical learning method. In addition, it is worth noting that our non-BRI index methods, which are all purely statistical, perform their best for nowcasts and backcasts. Indeed, they are MSFE-best in 15 cases out of 20 when considering only nowcasts and backcasts. Thus, the expert judgment associated with using the BRI index is not particularly useful for near-term forecasting, relative to existing methods of dimension reduction. However, when constructing predictions for $h = 1$ and $h = 2$, using the BRI index yields superior predictions in 19 cases out of 30.

Fifth, Fig. 1 plots the actual GDP growth, along with the monthly nowcasts obtained from the use of the BRI and SPCA selection methods. An examination of these plots indicates that DFM based on these dimension-reduction methods tends to predict turning points relatively well, outperforming the benchmark AR models particularly well during volatile episodes. This suggests that the selection of relevant predictors from a large datasets mitigates data noisiness that increases during periods of higher than normal volatility. Why? Perhaps because the correlations across a broad spectrum of variables increase during market downturns and volatile periods. This in turn magnifies the multicollinearity problem that characterizes the use of datasets where N is very large. Indeed, the efficacy of asymptotic theory associated with the use of principal components in time series contexts (see e.g. Bai & Ng, 2002, 2006, 2008) often relies on the assumption that the cross-correlations between the errors in factor models are not too large.

Finally, our results validate the findings of Boivin and Ng (2006), who suggest that correlations and data "noisiness" create a situation in which more data might not be

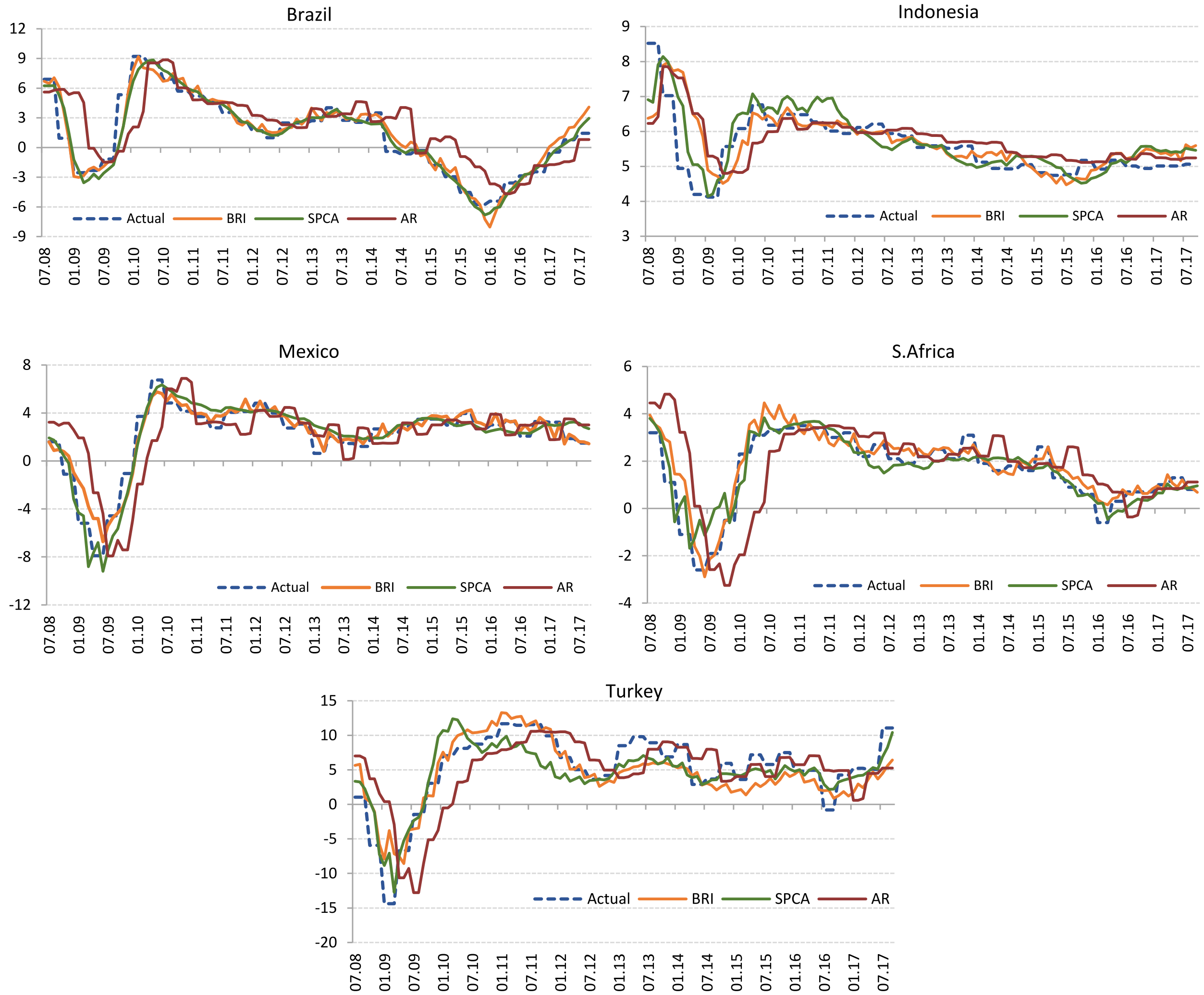

Fig. 1. Comparison of actual GDP growth rates with backcasts based on an AR benchmark and the BRI and SPCA dimension-reduction methods.

desirable. Indeed, we find that models that utilize expert judgment (BRI), as well as machine learning and shrinkage methods, yield very accurate forecasts when compared with both factor models that do not use dimension reduction (see entries denoted by ALL in Table 4), and benchmark autoregressive models. These findings confirm that imposing a sparse structure on the whole dataset is generally a useful step towards increasing predictive accuracy.

4.2. *Exploiting cross-country linkages when constructing diffusion indexes*

The objective of our second set of experiments is to provide a comprehensive empirical characterization of useful business cycle linkages between emerging markets using a dynamic factor model. In particular, we address the following question: Does taking cross-country business cycle factors into account lead to marginal gains in terms of predictive accuracy when analyzing emerging market economies? As was discussed in Section 3.3, we attempt to answer this question by estimating three additional factors, EM Global, EM Macro and EM Financial, and by utilizing these new diffusion indexes in our predictive modeling. Recall that the factors utilized in our first set of experiments were constructed using only local or "own-country" variables. Our EM diffusion indexes utilize data that is pooled across all EM economies. Before discussing the usefulness of these global common factors, it is of interest to investigate the correlations between the GDP growth rates across our EM economies. Table 2 reports the correlation coefficients among the growth rates of the five countries. These correlation coefficients range from 0.16 to 0.70, indicating that GDP growth rates are significantly correlated across countries. While this is not surprising, it does indicate that the global diffusion indexes that we are discussing in this section might indeed be useful for prediction.

Fig. 2 provides insights into the evolution of global and country-specific factors by plotting GDP growth against estimated common factors. As the table shows, both local and global common factors track GDP growth quite well, and therefore can be a good proxy for the GDP dynamics in these countries during the global financial crisis.

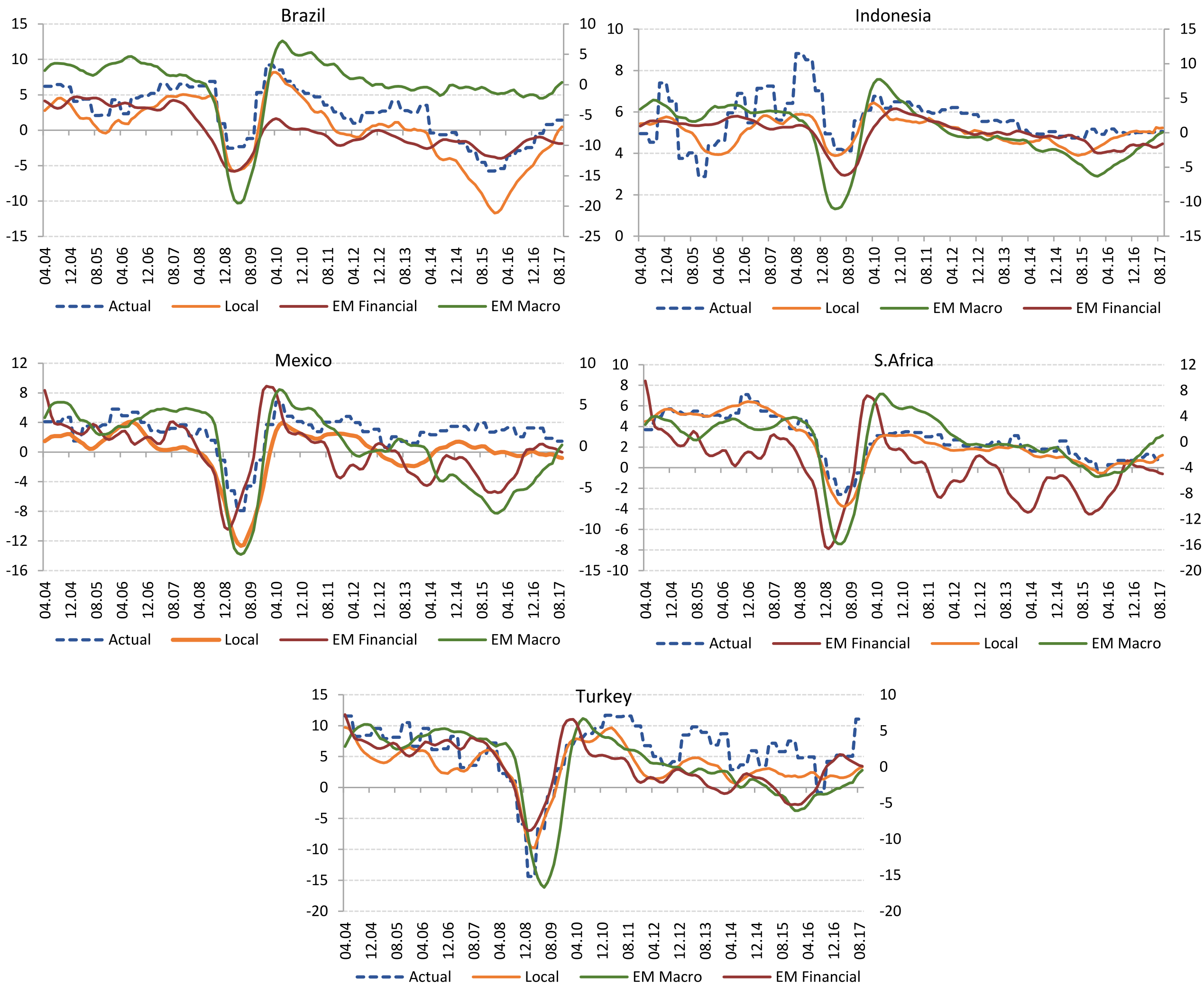

Fig. 2. GDP growth rates plotted against local (country-specific) and global (EM) diffusion indexes.

Tables 5–9 summarize the results of our second set of prediction experiments. As was discussed above, bold entries indicate the forecasting models that are the MSFE-best dimension-reduction methods. Thus, for BRI, when $h = 2$ and the month is "1", the EM Macro-Financial-AR model is MSFE-best. This means that our model that includes both global macroeconomic and global financial variables is superior to all other models when dimension reduction is carried out using the BRI index. Entries with a "GB" subscript added are the MSFE-best models for a given forecast horizon across *all* dimension-reduction methods. Thus, BRI is also the best dimension-reduction method across all six methods, including ALL, BRI, LASSO, ENET, LARS, and SPCA, for $h = 2$, when the month is "1". A summary of such MSFE-best models across dimension-reduction methods for each forecast horizon is given in Table A.2 of online appendix. An inspection of Tables 5–9 leads to a number of clearcut conclusions.

First, most of the entries in Tables 5–9 are below one, with only a limited number of exceptions, again indicating that dynamic factor model forecasts are more accurate than those constructed using our benchmark AR(SIC) models. The plethora of rejections of the null hypothesis of equal predictive accuracy when comparing our non-autoregressive type models with the AR(SIC) benchmark (see the many entries in the tables that are marked with either * or **, indicating DM test rejection) is further evidence of this finding. In addition, we again see that our factor models generally yield more accurate predictions as more information arrives, within each quarter.

Second, when comparing the "globally best" models across all forecast horizons, it is clear that the MSFE-best models are generally those that utilize dimension-reduction methods for selecting targeted predictors, and not those that are based on ALL or on the use of AR(SIC) models. This is further confirmation of our previous findings, as is the fact that BRI and SPCA win in 32 of 50 cases. Thus, data shrinkage and dimension-reduction methods are indeed useful.

Third, the models labeled "Local" and "Local-AR" in Tables 5–9 are not usually the MSFE-best models. Instead, models that include our global EM diffusion indexes are usually MSFE-best, for all forecast horizons, as well as

Table 5
MSFEs based on the use of different dimension-reduction and shrinkage methods with added global diffusion indexes: Brazil.

Full sample	Forecast ($h=2$)			Forecast ($h=1$)			Nowcast ($h=0$)			Backcast ($h=-1$)
	1	2	3	1	2	3	1	2	3	1
Local	1.01	0.92	0.89	0.83	0.73^{*}	0.67^{**}	0.52	0.38^{*}	0.29^{*}	0.42^{*}
EM Global	0.91	0.83	0.80	0.83	0.71^{*}	0.65^{**}	0.52^{*}	0.37^{*}	0.26^{*}	0.42^{*}
EM Macro	0.92	0.82	0.79	0.81	0.70^{*}	0.64^{**}	0.50^{*}	0.37^{*}	0.28^{*}	0.44^{*}
EM Macro-Financial	0.94	0.82	0.83	0.79	0.69^{*}	0.59^{**}	0.50^{*}	0.37^{*}	0.25^{*}	0.41^{*}
Local-AR	1.01	0.91	0.88	0.84	0.72^{*}	0.65^{*}	0.54	0.39^{*}	0.25^{*}	0.40^{*}
EM Global-AR	0.95	0.85	0.82	0.78^{*}	0.65^{**}	0.58^{**}	0.49^{*}	0.32^{*}	$\mathbf{0.21}^{*}$	$\mathbf{0.40}^{*}$
EM Macro-AR	**0.91**	$\mathbf{0.81}^{*}$	$\mathbf{0.78}^{**}$	$\mathbf{0.76}^{**}$	0.62^{**}	$\mathbf{0.56}^{**}$	0.48^{*}	0.31^{*}	0.22^{*}	0.40^{*}
EM Macro-Financial-AR	0.94	0.82	0.82	0.76^{*}	$\mathbf{0.61}^{**}$	0.57^{**}	$\mathbf{0.46}^{*}$	$\mathbf{0.28}^{**}$	0.24^{*}	0.42^{*}
BRI										
Local	0.94	0.84	0.83	0.81	0.66^{**}	0.63^{*}	0.53 *	0.36^{*}	0.34^{*}	0.55^{*}
EM Global	0.88	0.78^{*}	0.80^{*}	0.76^{**}	0.62^{**}	0.61^{**}	0.49^{*}	0.34^{*}	0.34^{*}	0.54^{*}
EM Macro	0.87	0.78^{*}	0.79^{*}	0.75^{**}	0.62^{**}	0.60^{**}	0.49^{*}	0.34^{*}	0.34^{*}	0.54^{*}
EM Macro-Financial	0.88	0.79	0.79	0.76^{**}	0.63^{**}	0.61^{**}	0.49^{*}	0.33^{*}	0.34^{*}	0.52^{*}
Local-AR	0.96	0.86	0.85	0.83	0.69^{**}	0.65^{*}	0.55	0.38^{*}	0.32^{*}	$\mathbf{0.47}^{*}$
EM Global-AR	0.90	0.79^{*}	0.80^{*}	0.77^{**}	0.63^{**}	0.61^{**}	0.50^{*}	0.33^{*}	0.31^{*}	0.48^{*}
EM Macro-AR	0.88	0.78^{*}	0.79^{*}	0.75^{**}	0.62^{**}	0.61^{**}	0.49^{*}	0.34^{*}	0.31^{*}	0.49^{*}
EM Macro-Financial-AR	$\mathbf{0.87}_{\mathbf{GB}}$	$\mathbf{0.77}_{\mathbf{GB}}$	**0.78**	$\mathbf{0.75}^{**}$	$\mathbf{0.61}^{**}$	$\mathbf{0.60}^{**}$	$\mathbf{0.49}^{*}$	$\mathbf{0.32}^{*}$	$\mathbf{0.30}^{*}$	0.48^{*}
LASSO										
Local	0.94	0.87	0.87	0.79^{*}	0.69^{*}	0.67^{**}	0.50^{*}	0.33^{*}	0.24^{*}	0.33^{*}
EM Global	0.91	0.83	0.83^{*}	0.75^{**}	0.64^{**}	0.61^{**}	0.47^{*}	0.29^{*}	0.20^{*}	0.35^{*}
EM Macro	0.90	0.82	0.82^{*}	0.75^{**}	0.64^{**}	0.61^{**}	0.47^{*}	0.30^{*}	0.20^{*}	0.34^{*}
EM Macro-Financial	0.93	0.85	0.86	0.78^{*}	0.67^{*}	0.64^{**}	0.50^{*}	0.31^{*}	0.19^{*}	$\mathbf{0.32}^{*}$
Local-AR	0.94	0.87	0.87	0.79^{*}	0.70^{*}	0.66^{**}	0.50^{*}	0.33^{*}	0.22^{*}	0.33^{*}
EM Global-AR	0.91	0.82	0.83^{*}	0.75^{**}	0.64^{**}	0.61^{**}	0.47^{*}	$\mathbf{0.28}^{**}$	0.20^{*}	0.36^{*}
EM Macro-AR	**0.90**	**0.82**	$\mathbf{0.82}^{*}$	$\mathbf{0.75}^{**}$	$\mathbf{0.64}^{**}$	$\mathbf{0.61}^{**}$	$\mathbf{0.47}^{*}$	0.30^{*}	0.20^{*}	0.35^{*}
EM Macro-Financial-AR	0.93	0.85	0.86	0.78^{*}	0.66^{*}	0.64^{**}	0.49^{*}	0.30^{*}	$\mathbf{0.19}_{\mathbf{GB}}^{*}$	0.33^{*}
ENET										
Local	0.97	0.90	0.90	0.83^{*}	0.71^{*}	0.69^{**}	0.54^{*}	0.35^{*}	0.27^{*}	0.35^{*}
EM Global	0.95	0.86	0.87	0.78^{**}	0.67^{**}	0.64^{**}	0.50^{*}	0.31^{*}	0.23^{*}	0.34^{*}
EM Macro	0.94	0.86	0.86	0.78^{**}	0.67^{**}	0.64^{**}	0.51^{*}	0.33^{*}	0.23^{*}	0.33^{*}
EM Macro-Financial	0.95	0.87	0.88	0.80^{*}	0.69^{**}	0.66^{**}	0.53^{*}	0.34^{*}	0.22^{*}	$\mathbf{0.31}_{\mathbf{GB}}^{*}$
Local-AR	0.96	0.89	0.89	0.82^{*}	0.70^{*}	0.67^{**}	0.53^{*}	0.34^{*}	0.25^{*}	0.34^{*}
EM Global-AR	0.94	0.85	0.85	0.77^{**}	0.65^{**}	0.62^{**}	$\mathbf{0.48}^{*}$	$\mathbf{0.30}^{**}$	0.22^{*}	0.35^{*}
EM Macro-AR	**0.93**	**0.84**	**0.84**	$\mathbf{0.77}^{**}$	$\mathbf{0.65}^{**}$	$\mathbf{0.62}^{**}$	0.50^{*}	0.31^{*}	0.22^{*}	0.35^{*}
EM Macro-Financial-AR	0.94	0.85	0.87	0.79^{*}	0.67^{**}	0.64^{**}	0.51^{*}	0.32^{*}	$\mathbf{0.21}^{*}$	0.32^{*}
LARS										
Local	0.97	0.90	0.85	0.81	0.70^{*}	0.63^{**}	0.50^{*}	0.35^{*}	0.24^{*}	0.44^{*}
EM Global	0.89	0.81^{*}	0.79^{*}	0.73^{**}	0.61^{**}	0.56^{**}	$\mathbf{0.44}_{\mathbf{GB}}^{*}$	0.27^{*}	0.22^{*}	0.44^{*}
EM Macro	0.88	0.80^{*}	0.78^{*}	0.74^{**}	0.62^{**}	$\mathbf{0.56}_{\mathbf{GB}}^{**}$	0.45^{*}	0.28^{*}	0.22^{*}	0.42^{*}
EM Macro-Financial	0.89	0.81^{*}	0.81^{*}	0.74^{**}	0.62^{**}	0.58^{**}	0.45^{*}	0.27^{*}	0.23^{*}	0.44^{*}
Local-AR	0.98	0.89	0.85	0.82	0.71^{*}	0.63^{**}	0.52^{*}	0.35^{*}	0.22^{*}	0.40^{*}
EM Global-AR	0.89	0.80^{*}	0.78^{**}	0.73^{**}	0.61^{**}	0.57^{**}	0.45^{*}	$\mathbf{0.27}_{\mathbf{GB}}^{**}$	0.20^{*}	0.42^{*}
EM Macro-AR	**0.88**	$\mathbf{0.79}^{*}$	$\mathbf{0.78}^{**}$	$\mathbf{0.73}_{\mathbf{GB}}^{**}$	$\mathbf{0.61}_{\mathbf{GB}}^{**}$	0.57^{**}	0.46^{*}	0.28^{*}	0.21^{*}	0.40^{*}
EM Macro-Financial-AR	0.89	0.80	0.83^{*}	0.74^{**}	0.61^{**}	0.60^{**}	0.45^{*}	0.27^{*}	$\mathbf{0.20}^{*}$	$\mathbf{0.40}^{*}$
SPCA										
Local	0.97	0.89	0.88	0.81	0.69^{*}	0.65^{*}	0.50^{*}	0.32^{*}	0.23^{*}	0.36^{*}
EM Global	**0.89**	**0.83**	$\mathbf{0.75}_{\mathbf{GB}}^{*}$	0.82	0.70^{*}	0.63^{*}	0.51^{*}	0.34^{*}	0.21^{*}	$\mathbf{0.35}^{*}$
EM Macro	0.97	0.88	0.87	0.81	0.69^{*}	0.63^{*}	0.51^{*}	0.33^{*}	0.21^{*}	0.36^{*}
EM Macro-Financial	0.99	0.92	0.90	0.81	0.71^{*}	0.65^{*}	0.51^{*}	0.35^{*}	0.23^{*}	0.37^{*}
Local-AR	0.96	0.87	0.87	0.79^{*}	0.66^{*}	0.62^{**}	$\mathbf{0.46}^{*}$	$\mathbf{0.28}^{**}$	0.21^{*}	0.40^{*}
EM Global-AR	0.95	0.87	0.87	0.79	0.67^{*}	0.63^{**}	0.47^{*}	0.29^{*}	0.20^{*}	0.38^{*}
EM Macro-AR	0.94	0.85	0.85	$\mathbf{0.78}^{*}$	$\mathbf{0.66}^{*}$	$\mathbf{0.62}^{**}$	0.47^{*}	0.29^{*}	$\mathbf{0.19}^{*}$	0.38^{*}
EM Macro-Financial-AR	0.99	0.90	0.90	0.81	0.69^{*}	0.65^{*}	0.49^{*}	0.32^{*}	0.23^{*}	0.41^{*}

Notes: See the notes to Table 4. The models utilized and reported on in Table 4 are augmented to include EM Global, EM Macro, and EM Financial diffusion indexes, which are constructed using global datasets, as discussed in Section 3.3. All models are listed in the first column of the table. Local, EM Global, EM Macro, and EM Macro-Financial correspond to Specifications 1–4 from Section 3.3, respectively; while the same models but with "-AR" appended are again Specifications 1–4, but with additional lagged dependent variables added as regressors. Entries in bold indicate the specifications that are "MSFE-best" for a particular predictor selection method, including ALL, BRI, LASSO, ENET, LARS, and SPCA, as discussed in Section 3.2. Entries in bold with a "GB" subscript are the MSFE-best models across all targeted predictor selection methods.

across all dimension-reduction methods. For example, in the case of Brazil, global EM diffusion indexes are included in the MSFE-best model for every prediction horizon and across every dimension-reduction method. The picture is

Table 6
MSFEs based on the use of different dimension-reduction and shrinkage methods with added global diffusion indexes: Indonesia.

Full sample	Forecast ($h = 2$)			Forecast ($h = 1$)			Nowcast ($h = 0$)			Backcast ($h = -1$)
	1	2	3	1	2	3	1	2	3	1
Local	1.23	1.47	1.16	1.21	1.36	0.97	1.20	1.12	0.73*	0.84
EM Global	1.70	1.91	1.73	1.42*	1.58	1.38	1.31	1.42	1.24	1.21
EM Macro	1.97	2.20	2.01	1.63*	1.70	1.47	1.34	1.49	1.26	1.23
EM Macro-Financial	2.13	2.35	1.90	1.61	1.78	1.25	1.30	1.50	1.13	1.45
Local-AR	1.17	1.43	1.11	1.11	1.28	**0.89**	1.08	**0.98**	**0.57***	**0.69***
EM Global-AR	1.57	1.84	1.48	1.29	1.48	1.11	1.04	1.17	0.84	0.83
EM Macro-AR	1.74	2.12	1.57	1.42	1.58	1.08	1.03	1.23	0.76	0.75
EM Macro-Financial-AR	2.11	2.37	1.82	1.57	1.75	1.09	1.09	1.31	0.78	0.93
BRI										
Local	**0.76$_{GB}$**	**0.75$_{GB}$**	**0.78$_{GB}$**	**0.72$_{GB}$***	**0.70$_{GB}$***	**0.69***	**0.82**	**0.81**	0.81	0.89
EM Global	1.18	1.23**	1.07	1.05	1.13	0.74	0.97	1.14	0.65	0.70
EM Macro	1.29	1.30	1.13	1.06	1.12	0.74	1.01	1.17	0.62	0.68
EM Macro-Financial	1.27	1.10	1.39	0.94	0.87	0.96	0.92	1.07	0.67	0.62*
Local-AR	1.00	1.00	0.99	0.99	0.98	0.94	0.95	0.91	0.84	0.83*
EM Global-AR	1.15	1.21	1.05	1.03	1.12	0.71	0.90	1.07	0.52*	0.55*
EM Macro-AR	1.26	1.30	1.11	1.06	1.13	0.73	0.97	1.12	**0.52$_{GB}$***	**0.54***
EM Macro-Financial-AR	1.60	1.51	1.62	1.13	1.06	1.17	0.87	0.88	0.82	0.74
LASSO										
Local	**0.85**	**0.84**	**0.87**	**0.80**	**0.80**	**0.75**	**0.94**	0.96	0.88	0.94
EM Global	1.89	2.06	1.84	1.46	1.53	1.19	1.06	1.08	0.78	0.78*
EM Macro	1.79	1.78	1.63	1.58	1.57	1.32	1.12	1.11	0.96	0.97
EM Macro-Financial	2.27	2.27	2.14	1.57	1.65	1.29	1.09	1.15	0.93	0.91
Local-AR	1.09	1.11	1.09	1.05	1.03	0.94	0.99	**0.95**	**0.68***	**0.62***
EM Global-AR	1.86	2.04	1.82	1.43	1.50	1.16	0.99	1.01	0.69*	0.67*
EM Macro-AR	2.33*	2.25	2.18	1.59	1.56	1.32	1.06	1.04	0.91	0.90*
EM Macro-Financial-AR	2.28*	2.30	2.17	1.57	1.65	1.32	1.07	1.12	0.93	0.90*
ENET										
Local	0.98	1.05	**0.95**	**0.86**	**0.89**	**0.78***	1.00	1.05	0.87	0.95
EM Global	1.80	1.82	1.77	1.48	1.45	1.33	1.28	1.21	1.10	1.06
EM Macro	2.37	2.07	2.44	1.70	1.54	1.57	1.35	1.26	1.26	1.26
EM Macro-Financial	2.16	1.82	2.19	1.50	1.37	1.33	1.19	1.14	1.02	1.10
Local-AR	**0.96**	**1.00**	0.96	1.01	1.00	0.88	**0.96**	**0.92**	**0.62***	**0.61***
EM Global-AR	1.66	1.71	1.67	1.37	1.35	1.22	1.14	1.06	0.89	0.81*
EM Macro-AR	2.21	1.90	2.24	1.55	1.37	1.37	1.11	1.00	0.94	0.94
EM Macro-Financial-AR	2.10	1.74	2.14	1.43	1.29	1.25	1.07	1.00	0.87	0.93
LARS										
Local	1.10	1.12	1.15	**0.98**	1.02	1.02	0.98	0.95	1.09	1.18
EM Global	1.35	1.58	1.47	1.08	1.19	1.08	0.82	0.86	0.72	0.58*
EM Macro	1.41	1.59	1.58	1.10	1.17	1.14	0.83	0.83	0.75	0.58*
EM Macro-Financial	1.56	1.45	1.48	1.17	1.09	1.01	0.91	0.85	0.79	0.72*
Local-AR	1.14	1.15	1.15	1.03	1.06	0.99	0.90	0.85	0.87	0.90
EM Global-AR	1.28	1.50	1.22	1.04	1.15	**0.91**	0.78	0.82	**0.56***	0.39*
EM Macro-AR	1.30	1.47	1.32	1.04	1.11	0.97	**0.78**	**0.77**	0.59	**0.38$_{GB}$****
EM Macro-Financial-AR	1.41	1.47	1.59	1.06	1.06	1.04	0.82	0.78	0.74	0.62*
SPCA										
Local	1.06	**0.99**	1.21	**0.80**	**0.74**	0.77	0.75	0.73	0.70	0.78
EM Global	1.66	1.51	1.58	1.09	1.09	0.92	0.85	0.87	0.76*	0.79*
EM Macro	1.96	1.41	2.24	1.08	0.89	1.16	0.81	0.81	0.99	0.99
EM Macro-Financial	2.13	2.64	2.29	1.33	1.81*	1.38	0.80	1.20	0.84	0.93
Local-AR	1.12	1.08	1.08	0.87	0.84	**0.68$_{GB}$***	0.74*	0.71*	**0.54***	**0.60***
EM Global-AR	1.68	1.52	1.58	1.07	1.06	0.90	0.77	0.78	0.66*	0.68*
EM Macro-AR	2.12	1.45	2.35	1.13	0.82	1.15	0.68*	**0.62$_{GB}$***	0.87	0.86
EM Macro-Financial-AR	2.27	2.67	2.57	1.34	1.80	1.63	**0.66$_{GB}$***	1.06	0.91	0.87

Note: See the notes to Table 5.

equally clear for Mexico, South Africa, and Turkey: global EM diffusion indexes yield substantial predictive gains, as "Local" and "Local-AR" are the MSFE-best models in only 12 of 30 cases, across all dimension-reduction methods for these countries. Interestingly, the same cannot be said for Indonesia, as "Local" and "Local-AR" are the MSFE-best models in six of 10 cases, across all dimension-reduction methods.[10] Given these findings, it should come as no

10 Note the low correlations between Indonesian GDP growth rates and the growth rates for Brazil, Mexico, South Africa and Turkey.

Table 7
MSFEs based on the use of different dimension-reduction and shrinkage methods with added global diffusion indexes: Mexico.

Full sample	Forecast ($h = 2$)			Forecast ($h = 1$)			Nowcast ($h = 0$)			Backcast ($h = -1$)
	1	2	3	1	2	3	1	2	3	1
Local	0.90	0.78	0.91	0.70*	0.58*	0.61*	0.49**	0.39**	0.37*	**0.42***
EM Global	**0.86**	**0.75**	**0.85**	0.72*	0.58*	0.67	0.47**	0.36**	0.43*	0.46*
EM Macro	1.02	0.91	1.08	0.68*	0.56*	0.62*	0.47**	0.37**	0.40*	0.46*
EM Macro-Financial	1.06	0.90	1.11	0.77	0.64*	0.74	0.50**	0.39**	0.50*	0.48*
Local-AR	0.90	0.79	0.92	0.69*	0.57*	**0.60***	0.48**	0.38**	**0.37***	0.44*
EM Global-AR	0.99	0.84	1.08	0.70*	0.57*	0.69	**0.45****	**0.33****	0.45*	0.58*
EM Macro-AR	0.91	0.79	0.94	**0.66***	**0.54**	0.62*	0.45**	0.35**	0.41*	0.59*
EM Macro-Financial-AR	1.14	0.95	1.19	0.79	0.64	0.79	0.45**	0.36**	0.54*	0.62
BRI										
Local	**0.56**$_{GB}$	**0.54**$_{GB}$	0.66	0.64*	0.54*	0.52**	0.45**	0.38**	**0.26**$_{GB}^{*}$	**0.49***
EM Global	0.72	0.64	0.63	0.61*	0.52*	0.45*	0.44**	0.37**	0.32*	0.54
EM Macro	0.70	0.62	0.61	**0.58**$_{GB}^{*}$	0.48*	**0.43**$_{GB}^{**}$	0.41**	0.34**	0.30*	0.54*
EM Macro-Financial	0.79	0.69	0.64	0.59*	**0.46**$_{GB}^{*}$	0.44**	**0.40**$_{GB}^{**}$	**0.30**$_{GB}^{**}$	0.34*	0.61
Local-AR	0.62	0.59	0.70	0.65*	0.55*	0.51**	0.45**	0.38**	0.26*	0.51*
EM Global-AR	0.71	0.63	0.60	0.58*	0.49*	0.44**	0.43**	0.37**	0.37*	0.57*
EM Macro-AR	0.69	0.61	**0.59**$_{GB}^{*}$	0.60*	0.51*	0.48*	0.47**	0.42**	0.46*	0.54*
EM Macro-Financial-AR	0.80	0.68	0.62	0.62*	0.48*	0.50**	0.47**	0.39**	0.51*	0.51*
LASSO										
Local	**0.88**	**0.77**	0.87	0.71*	0.60*	0.61*	0.52**	0.41**	0.34*	**0.38**$_{GB}^{*}$
EM Global	0.91	0.78	0.90	0.70*	0.59*	0.62*	0.50**	0.39**	0.35*	0.39*
EM Macro	0.89	0.77	**0.87**	0.69*	0.59*	0.60*	0.50**	0.40**	0.35*	0.40*
EM Macro-Financial	1.01	0.87	1.02	0.78	0.66*	0.70	0.54**	0.42**	0.44*	0.45*
Local-AR	0.88	0.77	0.88	0.69*	0.58*	0.59*	0.49**	0.38**	**0.30***	0.38*
EM Global-AR	0.92	0.79	0.94	0.69*	0.57*	0.62	**0.47****	0.36**	0.35*	0.42*
EM Macro-AR	0.89	0.77	0.88	**0.67***	**0.56***	**0.59***	0.48**	0.37**	0.34*	0.42*
EM Macro-Financial-AR	1.18	0.99	1.17	0.84	0.69	0.79	0.48**	**0.34****	0.51*	0.66
ENET										
Local	0.85	0.75	0.84	0.69*	0.59*	0.59*	0.52**	0.43**	0.36*	0.42*
EM Global	**0.80**	**0.72**	**0.77**	0.68*	0.58*	0.58*	0.50**	0.41**	0.35*	**0.41***
EM Macro	0.88	0.76	0.86	0.68*	0.59*	0.60*	0.51**	0.43**	0.39*	0.43*
EM Macro-Financial	0.97	0.83	0.98	0.75	0.64*	0.68	0.54**	0.44**	0.45*	0.45*
Local-AR	0.86	0.76	0.84	0.70*	0.60*	0.57*	0.52**	0.43**	**0.34***	0.41*
EM Global-AR	0.87	0.76	0.87	0.68*	**0.57***	**0.57***	0.50**	0.41**	0.35*	0.42*
EM Macro-AR	0.87	0.76	0.86	**0.67***	0.58*	0.58*	0.50**	0.42**	0.38*	0.44*
EM Macro-Financial-AR	1.08	0.91	1.15	0.77	0.65*	0.77	**0.47****	**0.37****	0.51*	0.53*
LARS										
Local	**0.83**	0.72	0.81	0.66*	0.55*	0.55*	0.48**	0.39**	0.31*	**0.39***
EM Global	0.84	**0.71**	**0.81**	**0.65***	**0.54***	**0.54***	0.48**	0.38**	0.31*	0.40*
EM Macro	0.85	0.72	0.82	0.65*	0.55*	0.55*	0.47**	0.39**	0.32*	0.41*
EM Macro-Financial	0.99	0.83	1.03	0.74*	0.59*	0.64*	0.46**	**0.32****	**0.27***	0.43*
Local-AR	0.85	0.73	0.82	0.67*	0.56*	0.55*	0.50**	0.39**	0.31*	0.40*
EM Global-AR	0.85	0.72	0.82	0.66*	0.55*	0.55*	0.49**	0.40**	0.34*	0.43*
EM Macro-AR	0.86	0.72	0.83	0.66*	0.55*	0.55*	0.48**	0.40**	0.35*	0.44*
EM Macro-Financial-AR	1.07	0.88	1.12	0.77	0.61*	0.75	**0.46****	0.32**	0.49*	0.51*
SPCA										
Local	1.09	0.92	1.17	0.74	0.58*	**0.73**	0.45**	0.33**	**0.51***	0.56*
EM Global	1.10	0.94	1.21	0.76	0.60*	0.76	0.48**	0.35**	0.52*	**0.55***
EM Macro	1.09	**0.91**	1.17	**0.72**	**0.56***	0.73	**0.44****	**0.33****	0.53*	0.59*
EM Macro-Financial	1.25	1.10	1.33	0.82	0.67*	0.86	0.51*	0.38**	0.61**	0.64
Local-AR	1.13	0.95	1.23	0.75	0.59*	0.77	0.46**	0.33**	0.54*	0.56*
EM Global-AR	1.15	1.00	1.30	0.75	0.59*	0.82	0.46**	0.35**	0.57**	0.61*
EM Macro-AR	1.12	0.95	1.22	0.73	0.57*	0.77	0.45**	0.35**	0.57*	0.61*
EM Macro-Financial-AR	1.59	1.47	1.47	0.96	0.83	0.95	0.52**	0.44**	0.69**	0.83

Note: See the notes to Table 5.

surprise that GDP growth rates are correlated contemporaneously with all of the diffusion indexes analyzed in this paper. Figure A.1 of online appendix plots correlation coefficients from simple regressions of the GDP growth in each country against our "Local" model, as well as against our global EM diffusion indexes. An inspection of this figure indicates that local diffusion indexes exhibit high correlations with GDP growth. However, the correlations with global EM indexes are also surprisingly high, supporting our finding that using both local and global indexes yields superior predictions for most countries, regardless of the variety of dimension reduction that is utilized.

Table 8
MSFEs based on the use of different dimension-reduction and shrinkage methods with added global diffusion indexes: South Africa.

Full sample	Forecast ($h = 2$)			Forecast ($h = 1$)			Nowcast ($h = 0$)			Backcast ($h = -1$)
	1	2	3	1	2	3	1	2	3	1
Local	0.91	0.85	0.83	0.81*	0.74**	0.66*	0.71**	0.64**	0.46**	0.43**
EM Global	0.93	0.86	0.86	0.81*	0.73**	0.63*	0.68**	0.60**	0.39**	0.37**
EM Macro	0.95	0.85	0.97	0.78*	0.70**	0.66*	0.63**	0.56**	0.40**	0.35**
EM Macro-Financial	1.02	0.90	1.20	0.79	0.67*	0.80	0.59**	0.48**	0.56**	0.51**
Local-AR	**0.88**	**0.82**	**0.78***	0.78*	0.71**	0.62*	0.65**	0.58**	0.41**	0.39**
EM Global-AR	0.89	0.82	0.81	0.77**	0.69**	**0.59***	0.62**	0.54**	**0.35****	**0.32****
EM Macro-AR	0.90	0.82	0.93	0.76**	0.66**	0.63*	0.60**	0.51**	0.39**	0.32**
EM Macro-Financial-AR	0.93	0.83	1.12	**0.76***	**0.63***	0.74	**0.57****	**0.46****	0.50**	0.42**
BRI										
Local	0.82	0.75	0.83	**0.66****	**0.58****	0.65**	0.43**	**0.37**$_{GB}^{**}$	0.41**	0.49**
EM Global	1.16	1.03	1.26	0.90	0.80	0.91	0.43**	0.44**	0.46**	0.55**
EM Macro	1.22	1.03	1.33	0.93	0.78	0.90	0.47**	0.42**	0.48**	0.62**
EM Macro-Financial	1.19	1.03	1.19	0.92	0.78	0.88	0.54**	0.45**	0.48**	0.63**
Local-AR	**0.79**	**0.74**	**0.77***	0.66*	0.60**	**0.61****	0.47**	0.41**	**0.37****	**0.41****
EM Global-AR	1.04	0.92	1.14	0.81	0.71*	0.81	**0.43****	0.41**	0.43**	0.52**
EM Macro-AR	1.10	0.93	1.18	0.84	0.69*	0.81	0.46**	0.39**	0.44**	0.57**
EM Macro-Financial-AR	1.15	0.98	1.15	0.89	0.74*	0.82	0.53**	0.43**	0.44**	0.60**
LASSO										
Local	0.82	0.82	0.70	0.74*	0.74	0.57*	0.64**	0.65*	0.47**	0.55**
EM Global	**0.79**	**0.73**	**0.67**	0.69**	0.62**	0.49*	0.59**	0.53**	0.33**	0.40**
EM Macro	0.84	0.73	0.78	**0.69****	**0.60****	0.52*	0.55**	0.50**	0.33**	0.40**
EM Macro-Financial	0.98	0.84	1.04	0.73*	0.61**	0.67*	**0.52****	**0.43****	0.45**	0.38**
Local-AR	0.80	0.80	0.68	0.74**	0.75	0.57*	0.64**	0.66	0.41**	0.52**
EM Global-AR	0.82	0.74	0.69	0.71**	0.63**	**0.49***	0.56**	0.49**	**0.29**$_{GB}^{**}$	0.31**
EM Macro-AR	0.83	0.75	0.77	0.70**	0.62**	0.52*	0.55**	0.47**	0.30**	**0.31**$_{GB}^{**}$
EM Macro-Financial-AR	0.89	0.81	1.01	0.72*	0.63**	0.66*	0.55**	0.46**	0.45**	0.34**
ENET										
Local	0.81	0.76	0.73	0.70**	0.64**	0.55*	0.63**	0.59**	0.46**	0.52**
EM Global	0.78	0.71	**0.67**	**0.68****	0.61**	0.51*	0.63**	0.59**	0.46**	0.52**
EM Macro	0.83	0.75	0.74	0.71**	0.62**	0.56*	0.61**	0.55**	0.46**	0.51**
EM Macro-Financial	**0.68**	**0.61**	0.86	0.73*	0.61**	0.75	0.58**	0.48**	0.59**	0.52**
Local-AR	0.83	0.78	0.73	0.73**	0.67**	0.56*	0.62**	0.56**	0.39**	0.39**
EM Global-AR	0.80	0.72	0.71	0.70**	0.61**	**0.49***	0.59**	0.51**	**0.35****	0.37**
EM Macro-AR	0.85	0.77	0.81	0.71**	0.62**	0.54*	0.58**	0.50**	0.36**	**0.36****
EM Macro-Financial-AR	0.71	0.63	1.04	0.69*	**0.56****	0.71*	**0.52****	**0.41****	0.53**	0.42**
LARS										
Local	0.86	0.80	0.78	0.75*	0.69**	0.61*	0.70**	0.67**	0.60**	0.68**
EM Global	**0.82**	0.77	0.73	0.71*	0.66**	0.59*	0.65**	0.62**	0.56**	0.64**
EM Macro	0.84	0.79	0.77	0.72*	0.66**	0.60*	0.64**	0.61**	0.56**	0.63**
EM Macro-Financial	0.94	0.84	0.90	0.78*	0.71**	0.66**	0.67**	0.62**	0.56**	0.58**
Local-AR	0.83	0.78	0.73	0.74**	0.67**	0.57*	0.62**	0.56**	0.41**	0.43**
EM Global-AR	0.83	**0.76**	**0.71**	**0.68****	0.62**	**0.51***	0.55**	0.50**	**0.34****	0.39**
EM Macro-AR	0.84	0.77	0.76	0.69**	**0.62****	0.53*	**0.54****	0.49**	0.35**	**0.38****
EM Macro-Financial-AR	0.87	0.79	0.87	0.71**	0.63**	0.59*	0.55**	**0.47****	0.40**	0.40**
SPCA										
Local	**0.58**$_{GB}$	0.51	**0.49**$_{GB}$	**0.43**$_{GB}^{**}$	0.43**	0.42*	**0.43**$_{GB}^{**}$	0.45**	0.51**	0.56**
EM Global	0.65	**0.51**$_{GB}$	0.64	0.51**	**0.40**$_{GB}^{**}$	**0.42**$_{GB}^{*}$	0.52**	0.57**	0.66	0.79**
EM Macro	0.79	0.66	0.90	0.62**	0.59**	0.74**	0.43**	0.40**	0.54	0.54**
EM Macro-Financial	0.88	0.70	1.06	0.74*	0.58**	0.67*	0.45**	0.39**	0.55**	0.68**
Local-AR	0.69	0.58	0.64	0.55**	0.51**	0.42*	0.45**	0.47**	**0.37****	0.39**
EM Global-AR	0.72	0.60	0.73	0.58**	0.46**	0.46**	0.46**	0.40**	0.39**	0.40**
EM Macro-AR	0.79	0.63	1.01	0.62**	0.47**	0.62**	0.46**	**0.38****	0.49**	**0.38****
EM Macro-Financial-AR	0.88	0.70	1.15	0.70**	0.51**	0.72*	0.51**	0.39**	0.54**	0.40**

Note: See the notes to Table 5.

In summary, we have strong evidence that global EM diffusion indexes have useful predictive content, suggesting that linkages across EM economies can be modeled using diffusion indexes, and are useful for predicting GDP growth in emerging market economies. Our findings are consistent with those of business cycle synchronization studies that have focused on the growing integration of emerging markets' economies, and find that such synchronization is likely to result in the transmission of economic shocks via trade and financial linkages.

Table 9
MSFEs based on the use of different dimension-reduction and shrinkage methods with added global diffusion indexes: Turkey.

Full sample	Forecast ($h=2$)			Forecast ($h=1$)			Nowcast ($h=0$)			Backcast ($h=-1$)
	1	2	3	1	2	3	1	2	3	1
Local	**0.82**	**0.75**	**0.80**	0.73**	0.65*	0.62*	0.63	0.60*	0.59*	0.70
EM Global	0.90	0.83	0.88	0.78	0.71*	0.69	0.69	0.65	*0.63	0.69
EM Macro	1.07	0.96	1.04	**0.61****	**0.55***	**0.59***	**0.56***	**0.56***	0.63	0.69
EM Macro-Financial	1.26	1.12	1.25	0.66**	0.62**	0.60*	0.57*	0.58*	0.64*	0.67
Local-AR	0.87	0.80	0.84	0.78*	0.70*	0.66*	0.64	0.59*	0.52*	0.58*
EM Global-AR	0.90	0.84	0.88	0.78	0.71*	0.68	0.64	0.58*	**0.48***	**0.48**$_{GB}^{*}$
EM Macro-AR	1.03	0.94	1.04	0.86	0.78	0.79	0.69	0.62	0.51*	0.49*
EM Macro-Financial-AR	1.33	1.14	1.30	1.06	0.93	0.94	0.80	0.70	0.55*	0.53*
BRI										
Local	**0.82**	**0.72**	0.74	**0.53**$_{GB}^{*}$	**0.53**$_{GB}^{*}$	0.72*	**0.60**	0.61*	0.65*	0.66*
EM Global	0.87	0.78	0.83	0.79*	0.73*	0.68*	0.69	0.68	0.62*	0.63*
EM Macro	0.89	0.77	0.82	0.79*	0.71*	0.65	0.67	0.67	0.62*	0.64
EM Macro-Financial	1.07	0.99	0.95	0.76*	0.70*	0.67	0.60	**0.53***	**0.42**$_{GB}^{*}$	0.59*
Local-AR	0.85	0.76	**0.72**	0.74*	0.69*	**0.62***	0.63	0.66*	0.60*	0.59*
EM Global-AR	0.90	0.81	0.85	0.81*	0.75*	0.68	0.68	0.67	0.56*	**0.56***
EM Macro-AR	0.95	0.82	0.86	0.82	0.73*	0.68	0.65	0.63	0.56*	0.59*
EM Macro-Financial-AR	1.02	0.87	0.86	0.87	0.78*	0.67	0.66	0.65	0.53*	0.58*
LASSO										
Local	0.90	0.82	0.88	0.77	0.68*	0.70	0.57	0.50*	0.51*	0.68
EM Global	0.92	0.87	0.88	0.78*	0.70*	0.68	0.60	0.54*	0.54*	0.68
EM Macro	**0.70**	**0.73**	0.73	0.76**	0.65**	0.73	0.61	0.55*	0.70	0.79
EM Macro-Financial	0.87	0.84	0.81	0.74*	0.69*	0.66*	0.54	0.49*	0.53*	0.75
Local-AR	0.91	0.84	0.89	0.79	0.70*	0.71	0.58	0.50*	0.47*	0.59*
EM Global-AR	0.94	0.88	0.90	0.79*	0.70*	0.69	0.56	0.49*	**0.45***	**0.56***
EM Macro-AR	0.71	0.75	**0.70**$_{GB}^{*}$	**0.65***	**0.64***	0.67	0.56	0.55	0.68	0.78
EM Macro-Financial-AR	0.87	0.85	0.83	0.73*	0.68*	**0.66***	**0.50***	**0.47***	0.48*	0.62*
ENET										
Local	0.74	**0.70**$_{GB}$	0.76	0.70**	0.64**	0.65*	0.68	0.67	0.70	0.83
EM Global	0.86	0.81	0.86	0.78*	0.72*	0.72	0.73	0.71	0.72	0.79
EM Macro	0.94	0.87	0.95	0.73**	0.67**	0.68*	0.66*	0.64*	0.59	0.75
EM Macro-Financial	0.71	1.02	0.89	0.65*	**0.64****	**0.56**$_{GB}^{*}$	**0.56**	**0.60**	0.73	0.68
Local-AR	0.81	0.77	0.81	0.75**	0.69*	0.68*	0.65	0.62*	0.59*	0.67*
EM Global-AR	0.88	0.82	0.87	0.77*	0.71*	0.70	0.65	0.61*	0.56*	0.61*
EM Macro-AR	0.94	0.88	0.96	0.80	0.74	0.76	0.66	0.61	0.57*	**0.61***
EM Macro-Financial-AR	**0.68**$_{GB}$	0.97	**0.75**	**0.64***	0.81	0.69**	0.56*	0.70**	**0.51***	0.74
LARS										
Local	1.24	0.88	0.87	0.87	**0.65**	**0.64**	0.46*	0.41*	0.48*	**0.53***
EM Global	1.38	1.06	1.19	0.94	0.71	0.79	0.50	**0.39**$_{GB}^{*}$	0.46*	0.64
EM Macro	1.34	1.05	1.14	0.91	0.71	0.76	0.50	0.39*	0.45*	0.64
EM Macro-Financial	1.45	1.09	1.17	0.97	0.75	0.78	0.51	0.40*	**0.43***	0.64
Local-AR	1.20	**0.88**	**0.86**	**0.84**	0.66	0.64	**0.45**$_{GB}^{*}$	0.41*	0.44*	0.70
EM Global-AR	1.36	1.05	1.18	0.93	0.72	0.79	0.52	0.40*	0.44*	0.58
EM Macro-AR	1.32	1.03	1.13	0.91	0.71	0.76	0.51*	0.40*	0.43*	0.57
EM Macro-Financial-AR	1.42	1.06	1.17	0.96	0.74	0.79	0.52*	0.41*	0.43*	0.64
SPCA										
Local	0.85	**0.72**	0.86	**0.70***	**0.60***	0.68	**0.54***	0.50*	0.66	0.79
EM Global	0.88	0.84	0.89	0.80	0.72*	0.75	0.62	0.55*	0.68	0.72
EM Macro	1.07	1.02	1.16	0.92	0.83	0.94	0.68	0.56	0.61*	0.64
EM Macro-Financial	1.06	1.19	1.25	0.92	0.97	1.00	0.68	0.62	0.63*	0.64
Local-AR	0.85	0.75	0.84	0.72*	0.63*	0.64*	0.56*	0.50*	0.55*	0.65*
EM Global-AR	**0.83**	0.80	**0.79**	0.74	0.68*	**0.63***	0.56*	**0.49***	**0.50***	**0.57***
EM Macro-AR	1.10	1.07	1.19	0.94	0.85	0.95	0.66	0.53	0.56*	0.60*
EM Macro-Financial-AR	1.10	1.26	1.35	0.94	1.00	1.03	0.66	0.60	0.58*	0.80

Note: See the notes to Table 5.

5. Conclusion

Dynamic factor models are used widely in the forecasting literature. However, relatively few studies have analyzed the usefulness of dimension reduction, machine learning, and shrinkage methods for selecting targeted predictors to be included in factor models. This paper compares the use of multiple such methods for the construction of "local" and "global" diffusion indexes, in the context of GDP growth prediction in emerging market (EM) economies. We find that dimension reduction matters. In particular, the so-called Bloomberg relevance index (BRI), which is related to crowd-sourcing coupled with expert opinion, as well as sparse principal component analysis, is particularly useful for selecting targeted predictors when constructing diffusion indexes. We also find that global diffusion indexes, which capture "spillover" effects among countries, are useful for nowcasting and forecasting EM GDP growth. In particular, exploiting the informational content of business cycle diffusion indexes based on macroeconomic and financial variables pooled across multiple economies leads to improved predictions of GDP growth, relative to the case where only "own-economy" variables are used for constructing diffusion indexes.

This paper is meant only as a starting point, as many questions remain unanswered. For example, it would be of interest to collect the Bloomberg relevance index in real-time, and to assess its usefulness for prediction. Currently, the index is available only as a point estimate of a variable's relevance. Collecting time series which measure the relevance of variables may yield further interesting insights into the usefulness of crowd-sourcing big data methods. In addition, our analysis focuses on emerging market economies. It remains to assess how one might utilize the linkages between developed and emerging markets when predicting economic variables for EM economies.

Acknowledgments

The authors wish to thank the editor, Michael McCracken, two anonymous referees, Hyun Hak Kim, Mingmian Cheng, Eric Ghysels, Massimiliano Marcellino, Christian Schumacher, Marta Banbura, Domenico Giannone and Xiye Yang for useful discussions and comments on earlier versions of this paper. We also thank Michele Modugno and Matteo Luciani for sharing their Matlab codes.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2018.10.008.

References

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

Bai, J., & Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, *74*, 1133–1150.

Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*, 304–317.

Banbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-casting and the real-time data flow. In G. Elliot, & A. Timmermann (Eds.), *Handbook of economic forecasting, vol. 2*. Elsevier: Amsterdam.

Banbura, M., & Rünstler, G. (2011). A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, *27*, 333–346.

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

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

Bragoli, D. (2017). Now-casting the Japanese economy. *International Journal of Forecasting*, *33*, 390–402.

Bragoli, D., Metelli, L., & Modugno, M. (2014). The importance of updating: evidence from a Brazilian nowcasting model. Board of Governors of Federal Reserve System, Working Paper No. 2014-94.

Bulligan, G., Marcellino, M., & Venditti, F. (2015). Forecasting economic activity with targeted predictors. *International Journal of Forecasting*, *31*, 188–206.

Caruso, A. (2017). Nowcasting with the help of foreign indicators: the case of Mexico. *Economic Modelling*, *69*, 160–168.

Corradi, V., & Swanson, N. R. (2006). Predictive density evaluation. In G. Elliot, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting, vol. 1* (pp. 197–284). Elsevier: Amsterdam.

Corradi, V., & Swanson, N. R. (2007). Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes. *International Economic Review*, *48*, 67–109.

Dalhaus, T., Guénette, J. D., & Vasishtha, G. (2017). Nowcasting BRIC+M in real time. *International Journal of Forecasting*, *33*, 915–935.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*, 134–144.

Doz, C., Giannone, D., & Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, *164*, 188–205.

Doz, C., Giannone, D., & Reichlin, L. (2012). A quasi–maximum likelihood approach for large, approximate dynamic factor models. *Review of Economics and Statistics*, *94*, 1014–1024.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32*, 407–499.

Eickmeier, S., & Ng, T. (2011). Forecasting national activity using lots of international predictors: An application to New Zealand. *International Journal of Forecasting*, *27*, 496–511.

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

Foroni, C., & Marcellino, M. (2014). A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates. *International Journal of Forecasting*, *30*, 554–568.

Fu, W. J. (1998). Penalized regressions: the bridge versus the lasso. *Journal of Computational and Graphical Statistics*, *7*, 397–416.

Garcia, M. G., Medeiros, M. C., & Vasconcelos, G. F. (2017). Real-time inflation forecasting with high-dimensional models: The case of Brazil. *International Journal of Forecasting*, *33*, 679–693.

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

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*, 453–497.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised learning. In T. Hastie, R. Tibshirani, & J. Friedman (Eds.), *The elements of statistical learning* (2nd ed.). (pp. 485–585). Springer: New York.

Hindrayanto, I., Koopman, S. J., & de Winter, J. (2016). Forecasting and nowcasting economic growth in the euro area using factor models. *International Journal of Forecasting*, *32*, 1284–1305.

Kim, H. H., & Swanson, N. R. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*, *178*, 352–367.

Kim, H. H., & Swanson, N. R. (2018). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting*, *34*, 339–354.

Kuzin, V., Marcellino, M., & Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: nowcasting GDP in the euro area. *International Journal of Forecasting*, *27*, 529–542.

Kuzin, V., Marcellino, M., & Schumacher, C. (2013). Pooling versus model selection for nowcasting GDP with many predictors: empirical evidence for six industrialized countries. *Journal of Applied Econometrics*, *28*, 392–411.

Li, J., & Chen, W. (2014). Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting*, *30*, 996–1015.

Luciani, M., Pundit, M., Ramayandi, A., & Veronese, G. (2017). Nowcasting Indonesia. *Empirical Economics*, *52*, 1–23.

Luciani, M., & Ricci, L. (2014). Nowcasting Norway. *International Journal of Central Banking*, *10*(4), 215–248.

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

McCracken, M. W. (2000). Robust out-of-sample inference. *Journal of Econometrics*, *99*, 195–223.
Modugno, M., Soybilgen, B., & Yazgan, E. (2016). Nowcasting Turkish GDP and news decomposition. *International Journal of Forecasting*, *32*, 1369–1384.
Naikal, N., Yang, A. Y., & Sastry, S. S. (2011). Informative feature selection for object recognition via sparse PCA. In *2011 IEEE international conference on computer vision (ICCV)* (pp. 818–825).
Schumacher, C. (2007). Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting*, *26*, 271–302.
Schumacher, C. (2010). Factor forecasting using international targeted predictors: the case of German GDP. *Economics Letters*, *107*, 95–98.
Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*, 147–162.
Stock, J. H., & Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, *30*, 481–493.
Swanson, N. R. (2016). Comment on: In sample inference and forecasting in misspecified factor models. *Journal of Business & Economic Statistics*, *34*, 348–353.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B.*, *58*, 267–288.
Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *68*(1), 49–67.
Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*, 1418–1429.
Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B.*, *67*, 301–320.
Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, *15*, 262–286.

Oguzhan Cepni is a Director at the Central Bank of the Republic of Turkey. His primary research interests include applied econometrics, advanced analytics, machine learning and factor models.

Ibrahim Ethem Guney has a Ph.D. in Banking & Finance from University of Zurich and Swiss Finance Institute. He is currently Executive Director of Strategy and Corporate Governance Department at the Central Bank of the Republic of Turkey. His primary research interests include monetary policy, asset pricing, applied financial economics, mathematical finance and big data applications.

Norman R. Swanson was educated at the University of Waterloo and the University of California, San Diego. He is currently Professor and Director of Graduate Studies in the Economics Department at Rutgers University. He has held previous positions at Pennsylvania State University (1994–1999), Texas A&M University (1999–2001), and Purdue University (2001–2002). His primary research interests include econometric theory, financial and macro econometrics, time series analysis, and forecasting. He is a fellow of the Journal of Econometrics, he is currently an associate editor of various econometrics journals, and he has served as guest editor for journals ranging from Journal of Econometrics to Journal of Business and Economic Statistics. Swanson has recent scholarly publications in leading economics and statistics journals including Econometrica, Journal of Econometrics, Review of Economics and Statistics, Journal of Business and Economic Statistics, and the Journal of the American Statistical Association, among others; and is or has acted as ongoing visiting scholar to various central banks including The Federal Reserve Bank of Philadelphia and the Bank of Canada.