

WORKING PAPER

Forecasting GDP with global components. This time is different

NORGES BANK
RESEARCH

05 | 2015

AUTHORS:

HILDE C. BJØRNLAND
FRANCESCO RAVAZZOLO
LEIF ANDERS THORSRUD



NORGES BANK

Working papers fra Norges Bank, fra 1992/1 til 2009/2 kan bestilles over e-post:

servicesenter@norges-bank.no

Fra 1999 og senere er publikasjonene tilgjengelige på www.norges-bank.no

Working papers inneholder forskningsarbeider og utredninger som vanligvis ikke har fått sin endelige form. Hensikten er blant annet at forfatteren kan motta kommentarer fra kolleger og andre interesserte. Synspunkter og konklusjoner i arbeidene står for forfatterens regning.

Working papers from Norges Bank, from 1992/1 to 2009/2 can be ordered by e-mail

servicesenter@norges-bank.no

Working papers from 1999 onwards are available on www.norges-bank.no

Norges Bank's working papers present research projects and reports (not usually in their final form) and are intended inter alia to enable the author to benefit from the comments of colleagues and other interested parties. Views and conclusions expressed in working papers are the responsibility of the authors alone.

ISSN 1502-8143 (online)

ISBN 978-82-7553-857-2 (online)

Forecasting GDP with global components. This time is different*

Hilde C. Bjørnland[†] Francesco Ravazzolo[‡]
Leif Anders Thorsrud[§]

March 9, 2015

Abstract

A long strand of literature has shown that the world has become more global. Yet, the recent Great Global Recession turned out to be hard to predict, with forecasters across the world committing large forecast errors. We examine whether knowledge of in-sample co-movement across countries could have been used in a more systematic way to improve forecast accuracy at the national level. In particular, we ask if a model with common international business cycle factors forecasts better than the purely domestic alternative? To answer this question we employ a Dynamic Factor Model (DFM) and run an out-of-sample forecasting experiment. Our results show that exploiting the informational content in a common global business cycle factor improves forecasting accuracy in terms of both point and density forecast evaluation across a large panel of countries. In line with much reported in-sample evidence, we also document that the Great Recession has a huge impact on this result. The event causes a clear preference shift towards the model including a common global factor. Similar shifts are not observed earlier in the evaluation sample. However, this time is different also in other respects. On longer forecasting horizons the performance of the DFM deteriorates substantially in the aftermath of the Great Recession. This indicates that the recession shock itself was felt globally, but that the recovery phase has been very different across countries.

JEL-codes: C11, C53, C55, F17

Keywords: Bayesian Dynamic Factor Model (BDFM), forecasting, model uncertainty and global factors

*This working paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank. The authors would like to thank seminar and conference participants at PBC School of Finance at Tsinghua University, CAMA at Australia National University, the 22nd Symposium of the Society for Nonlinear Dynamics and Econometrics in New York, the CEF 2014 annual conference in Oslo and Norges Bank for valuable comments. This paper is part of the research activities at the Centre for Applied Macro and Petroleum economics (CAMP) at the BI Norwegian Business School.

[†]BI Norwegian Business School and Norges Bank. Email: hilde.c.bjornland@bi.no

[‡]Norges Bank and BI Norwegian Business School. Email: francesco.ravazzolo@norges-bank.no

[§]BI Norwegian Business School. Email: leif.a.thorsrud@bi.no

1 Introduction

A long strand of literature has shown that the co-movement of aggregate activity across a large number of countries can be reasonably well explained by one (or a few) common business cycle factors. For example, in [Kose et al. \(2003\)](#) up to 35 percent of the variance in domestic GDP across G7 countries is attributed to one common international business cycle factor.¹ [Imbs \(2010\)](#) provides further evidence. Focusing on the distribution of business cycles, he shows that the Great Recession was a true global recession, the first in decades.

Knowledge of in-sample co-movement across countries, in particular during recessions, does not necessarily imply predictability. As is now well known, the Great Recession (2007-2009) turned out to be hard to predict, with private and public sector forecasters across the world committing large forecast errors relative to their historical performance, c.f. [Stockton \(2012\)](#) and [Alessi et al. \(2014\)](#) for a discussion of the forecast performance of the Bank of England, Federal Reserve Bank of New York and the European Central Bank. A common explanation for this forecast failure has been the high level of uncertainty at the time. Neither the size or timing of the shocks that hit the global economy had been foreseen. Nor was the extent of the propagation of these shocks into economic activity. Consistent with this, many recent papers have shown that forecast errors typically rise dramatically in severe recessions when uncertainty is high, c.f., [Baker et al. \(2013\)](#), and most notably during the Great Recession, see [Jurado et al. \(2013\)](#).²

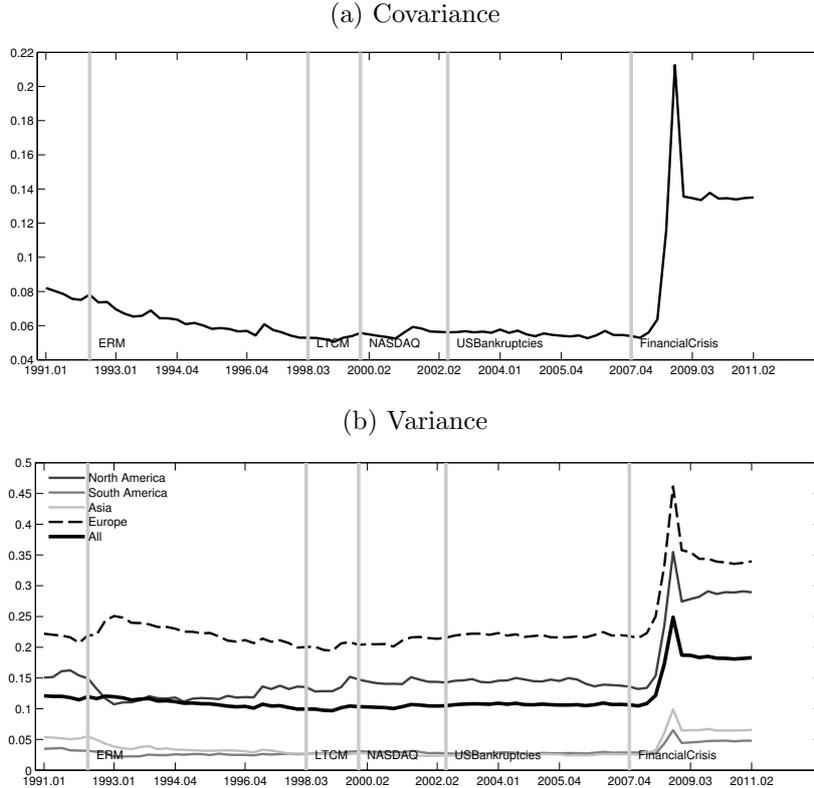
Our purpose is to examine whether knowledge of in-sample co-movement across countries could have been used in a more systematic way to improve forecast accuracy at the national level. As claimed by, e.g., [Ashley et al. \(1980\)](#), in-sample inference without out-of-sample verification is likely to be spurious, with an out-of-sample approach inherently involving less over-fitting. Hence, an out-of sample evaluation seems a natural next step in light of the in-sample evidence of co-movement across countries (reported in, e.g., [Imbs \(2010\)](#)), and the massive forecast failures across the world prior to the Great Recession. This paper therefore asks: Does a model with common international business cycle factors forecast better than the purely domestic alternative, i.e., a simple autoregressive process?

To examine whether common international components add value in terms of forecasting, we specify a Dynamic Factor Model (DFM) widely used for studying international business cycle synchronization. Our dataset contains quarterly real GDP growth from 1978 to 2011 for 33 countries across the world, broadly covering 4 geographical regions and both developed and emerging economies. The out-of-sample forecasting experiment starts in 1991:Q1. For each new vintage of data the DFM is re-estimated and forecasts produced, generating a total of 82 out-of-sample forecast observations for each country. The factor model forecasts are compared to forecasts produced by a simple autoregressive process. In our experiment this is the natural benchmark model,

¹Corroborated theoretical and empirical findings are reported in, e.g., [Ambler et al. \(2002\)](#), [Stock and Watson \(2005\)](#), [Baxter and Kouparitsas \(2005\)](#), [Backus and Kehoe \(1992\)](#), [Backus et al. \(1995\)](#) and [Engel and Wang \(2011\)](#).

²More generally, GDP is often found to be hard to predict, and it has been difficult to beat an $AR(p)$ model or a random walk, see, e.g., [Stock and Watson \(1999, 2002\)](#). Furthermore, the predictability of more complex models relative to naive forecasts seems to have declined since the 1980s, see [D'Agostino et al. \(2006\)](#).

Figure 1: In-sample evidence



Note: Figure 1a (1b) reports the recursively estimated covariance (variance) as implied by the Dynamic Factor Model and one common global business cycle factor, see Section 2 for a detailed description of the model. The covariance (variance) is reported as the average of all elements below (on) the diagonal in the covariance matrix at each point in time. The variance is reported as a fraction of the total variance explained by the model. In addition, Figure 1b reports the average within different regional clusters.

as the DFM we employ collapses to an autoregressive process for each country if the number of common factors is 0, i.e., if the common factors are irrelevant.³ We evaluate predictability in terms of both point and density forecasts across time, employing mean squared errors (MSE) and continuous ranked probability score (CRPS) scoring functions. In particular, we examine whether predictability increased in the recent financial crisis relative to previous recessions.

The DFM we employ is similar to the factor models used in other business cycle synchronization studies. For this reason, we would expect the model to also confirm the earlier in-sample evidence alluded to above - and it does. In line with the results reported in Imbs (2010), the Great Recession had a huge impact on business cycle synchronization, increasing the covariance across countries explained by one common business cycle factor considerably, see Figure 1a. Much of this increased covariance can be attributed to an increase in overall volatility in the period around the Great

³In a pure forecasting horse-race, other benchmark models would very likely be harder to outperform. In addition, when assessing predictability across 33 countries, the best benchmark model would almost surely vary considerably across countries.

Recession. This is seen in Figure 1b, which reports the variance explained attributed to the common business cycle factor. Overall, the variance explained attributed to the common factor is not large, but for countries in Europe and North America it is substantial.⁴ As such, the figure also highlights another important feature found in the more recent business cycle literature, namely that the world is not enough: When explaining business cycle synchronization, common regional factors seem to matter more and more relative to one common global factor, see, e.g., [Crucini et al. \(2011\)](#), [Mumtaz et al. \(2011\)](#) and [Thorsrud \(2013\)](#). In this paper we will address the uncertainty in the number of factors by using the out-of-sample forecasting performance as a measure of model fit.

Our out-of-sample forecasting experiment delivers the following results: First, exploiting the information content in a common global business cycle factor improves forecasting accuracy in terms of both point and density forecast evaluation across a large panel of countries. In particular we find that the forecasts produced by the standard DFM on average (across countries) adds marginal predictive power to the natural benchmark, an autoregressive process for each individual country.

Second, in line with the in-sample evidence reported above, we also document that the Great Recession has a huge impact on this result. Irrespective of which loss function we use, the event causes a clear preference shift towards the model including a common global factor. Similar shifts are not observed earlier in the evaluation sample. This is in particular interesting in light of the recent evidence of heightened uncertainty and increased forecast errors during deep recessions, c.f. [Baker et al. \(2013\)](#) and [Jurado et al. \(2013\)](#). In our out-of sample forecasting experiment the information content in the common global component now works to reduce the forecast errors (relative to the AR(1)) and hence increase forecasting accuracy.

However, this time is different also in other respects. On longer forecasting horizons the performance of the DFM deteriorates substantially in the aftermath of the Great Recession. To the extent that forecast errors are a good proxy for uncertainty, this indicates that during the recession, information about common movements could have worked to reduce uncertainty across countries, while during the recovery phase use of such knowledge would in fact have increased it. One potential reason for this might be uncertainty related to economic policies in the US and Europe in particular, see, e.g., [Baker et al. \(2013\)](#). Augmenting the DFM with regional factors (including, e.g., an Asia-specific business cycle factor) alleviates this uncertainty somewhat: It improves the short-term forecasting performance of the model further, and gives out-of-sample support to the in-sample studies advocating the importance of regional business cycle factors. Finally, incorporating uncertainty regarding the true number of factors by employing a forecast combination approach confirms that regional factors are important.

The rest of the paper is organized as follows: In Section 2 we describe the DFM, the data, and the out-of-sample forecasting experiment. In Section 3 we report our results. We start with our main evaluation using a DFM with one global component only. We then report the results for the augmented model, including regional factors, and the results for the model combination experiment. Section 4 concludes.

⁴Even for countries in these regions the numbers are somewhat smaller than those reported in [Kose et al. \(2003\)](#). One reason for this is likely the fact that we look at quarterly growth rates, while they use yearly numbers that contain much less noise.

2 Model, estimation and evaluation under model uncertainty

As alluded to above, we entertain a Dynamic Factor Model (DFM) frequently employed in the business cycle synchronization literature. This model is particularly useful in a data rich environment such as ours, where common latent factors and shocks are assumed to drive the co-movements across a large cross section of countries.

The DFM is given by equations 1 and 2:

$$y_t = \lambda_0 f_t + \dots + \lambda_s f_{t-s} + \epsilon_t \quad (1)$$

where the $N \times 1$ vector y_t represents the observables at time t . λ_j is a $N \times q$ matrix with dynamic factor loadings for $j = 0, 1, \dots, s$, and s denotes the number of lags used for the dynamic factors f_t . Lastly, ϵ_t is an $N \times 1$ vector of idiosyncratic errors.

The dynamic factors follow a VAR(h) process:

$$f_t = \phi_1 f_{t-1} + \dots + \phi_h f_{t-h} + u_t \quad (2)$$

where u_t is a $q \times 1$ vector of VAR(h) residuals. The idiosyncratic and VAR(h) residuals are assumed to be independent:

$$\begin{bmatrix} \epsilon_t \\ u_t \end{bmatrix} \sim i.i.d.N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R & 0 \\ 0 & Q \end{bmatrix} \right) \quad (3)$$

Further, in our application R is assumed to be diagonal.

We consider the case where $\epsilon_{t,i}$, for $i = 1, \dots, N$, follows independent AR(1) processes:

$$\epsilon_{t,i} = \rho_{1,i} \epsilon_{t-1,i} + \dots + \rho_{l,i} \epsilon_{t-l,i} + \omega_{t,i} \quad (4)$$

where l denotes the number of lags, and $\omega_{t,i}$ is the AR(1) residuals with $\omega_{t,i} \sim i.i.d.N(0, \sigma_i^2)$. I.e.:

$$R = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \sigma_N^2 \end{bmatrix}, \quad (5)$$

To separately identify the factors and the loadings, and to be able to give the factors an economic interpretation, we enforce the following identification restrictions on equation 1:

$$\lambda_0 = \begin{bmatrix} \lambda_{0,1} \\ \lambda_{0,2} \end{bmatrix} \quad (6)$$

where $\lambda_{0,1}$ is a $q \times q$ identity matrix, and $\lambda_{0,2}$ is left unrestricted. As shown in [Bai and Ng \(2010\)](#) and [Bai and Wang \(2012\)](#), these restrictions uniquely identify the dynamic factors and the loadings, but leave the VAR(h) dynamics for the factors completely unrestricted.

2.1 Data

Our data set is composed of real Gross Domestic Product (GDP) for a large cross section, covering 33 countries. The data are collected from the GVAR ‘2011 Vintage’ data set constructed by Gang Zhang, Ambrogio Cesa Bianchi, and Alessandro Rebucci at the Inter-American Development Bank. The cross section of GDP data covers the countries: Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, China, Finland, France, Germany, India, Indonesia, Italy, Japan, Korea, Malaysia, Mexico, The Netherlands, New Zealand, Norway, Peru, Philippines, Saudi Arabia, Singapore, South Africa, Spain, Sweden, Switzerland, Thailand, Turkey, the United Kingdom, and the United States.

The data span the time period 1978:Q1 to 2011:Q2. Prior to estimation, all data are made stationary by taking the logarithmic difference.

Table 3, in Appendix B, reports descriptive statistics for GDP for the 33 countries under study. As seen in the table there is large, and well known, heterogeneity across countries. Mean GDP growth is higher (above median) for most emerging Asian countries and some commodity exporters (i.e., Australia, Brazil, Mexico, Norway and Peru), and lower for G8 economies. Moreover, volatility is higher for small open economies (i.e., Norway and Sweden) and for most Asian and South American countries and lower for G8 and most other European countries.

2.2 Specifications and estimation

In the baseline scenario we specify a factor model including only one global business cycle factor. That is, we set $q = 1$ and identify it as a global business cycle factor by letting US GDP be the first variable in the observable y_t vector. Accordingly, the global business cycle factor loads with one on US GDP growth. As described in greater detail in below, we later expand the model to include up to $q = 4$ regional business cycle factors.

For all specifications we let the number of lags in the transition equation equal two ($h = 2$), the number of lags of the dynamic factors equal zero ($s = 0$), and the number of lags for the idiosyncratic auto-regressions equal one ($l = 1$).

Let $\tilde{y}_T = [y_1, \dots, y_T]'$ and $\tilde{f}_T = [f_1, \dots, f_T]'$, and define $H = [\lambda_0, \dots, \lambda_s]$, $\beta = [\phi_1, \dots, \phi_h]$, Q , R , and $p_i = [\rho_{1,i}, \dots, \rho_{l,i}]$ for $i = 1, \dots, N$, as the model’s hyper-parameters. Inference on the unknown states and hyper-parameters is based on Bayesian estimation of the state space model and Gibbs simulation, where the following three steps are iterated until convergence is achieved:

Step 1: Conditional on the data (\tilde{y}_T) and all the parameters of the model, generate \tilde{f}_T

Step 2: Conditional on \tilde{f}_T , generate β and Q

Step 3: Conditional on \tilde{f}_T , and data for the i -th variable ($\tilde{y}_{T,i}$), generate H_i , R_i and p_i for $i = 1, \dots, N$

In Appendix A we describe each step in more detail and document the employed prior specifications. We simulate the model using a total of 10000 iterations. 5000 draws are used as burn-in, and only every 5th iteration is stored and used for inference.⁵

⁵Standard MCMC convergence tests, conducted on the model estimated on the full sample, confirm that the Gibbs sampler converges to the posterior distribution. Convergence statistics can be reported

2.3 Out-of-sample forecasting and evaluation

In constructing density and point forecasts, we use a recursive forecasting scheme, expanding the model estimation sample as forecasting moves forward in time. We focus on the full sample period 1990:Q1-2011:Q2, a pre-financial crisis sample period 1990:Q1-2007:Q3, and the financial crisis period 2007:Q4-2011:Q2. The section proceeds by detailing our approaches for scoring and comparing forecasts.

We consider several evaluation statistics for point and density forecasts previously proposed in literature. We compare point forecasts in terms of Root Mean Square Prediction Errors (RMSPE):

$$RMSPE_k = \sqrt{\frac{1}{t^*} \sum_{\tau=\underline{t}}^{\bar{t}} e_{k,\tau+h}} \quad (7)$$

where $t^* = \bar{t} - \underline{t} + h$, \bar{t} and \underline{t} denote the beginning and end of the evaluation period, and $e_{k,\tau+h}$ is the square prediction error associated to the forecast made by model k at time τ for the observation $y_{\tau+h}$.

Following [Welch and Goyal \(2008\)](#) we investigate how square prediction varies over time by a graphical inspection of the Cumulative Squared Prediction Error Difference (CSPED):

$$CSPED_{k,\tau+1} = \sum_{s=\underline{t}}^t \hat{f}_{k,s+h}, \quad (8)$$

where $\hat{f}_{k,\tau+1} = e_{AR,\tau+h} - e_{k,\tau+h}$. Increases in $CSPED_{k,\tau+h}$ indicate that the alternative model compared to the benchmark (AR model) predicts better at out-of-sample observation $\tau + h$.

Following [Gneiting and Ranjan \(2011\)](#), [Groen et al. \(2012\)](#) and [Ravazzolo and Vahey \(2013\)](#) for applications to inflation density forecasts, and [Clark and Ravazzolo \(2013\)](#) for a larger set of macro variables, we evaluate density forecasts based on the continuous rank probability score (CRPS). The CRPS for the model k measures the average absolute distance between the empirical cumulative distribution function (CDF) of $y_{\tau+h}$, which is simply a step function in $y_{\tau+h}$, and the empirical CDF that is associated with model k 's predictive density:

$$CRPS_{k,\tau+h} = \int (F(z) - \mathbb{I}_{[y_{\tau+h}, +\infty)}(z))^2 dz \quad (9)$$

$$= \mathbb{E}_t |\tilde{y}_{k,t+\tau} - y_{\tau+h}| - \frac{1}{2} \mathbb{E}_t |\tilde{y}_{k,\tau+h} - \tilde{y}'_{k,\tau+h}|, \quad (10)$$

where F is the CDF from the predictive density $p(\tilde{y}_{k,\tau+h}|y_{1:t})$ of model k and $\tilde{y}_{k,\tau+h}$ and $\tilde{y}'_{k,\tau+h}$ are independent random variables with common sampling density equal to the posterior predictive density $p(\tilde{y}_{k,\tau+h}|y_{1:t})$. Smaller CRPS values imply higher precisions and we report in tables the average $CRPS_k$ for each model k . [Gneiting and Raftery \(2007\)](#) discuss the properties of the CRPS, including that it is a strictly proper scoring rule that can be related to Bayes factors and to cross-validation, and can also be used for evaluation of some areas (quantiles) of interest of the predictive density.

on request.

As for the RMSPE analysis, we also compute Cumulative CRPS (CRPSD), and plot them.

To provide a rough gauge of whether the differences in forecast accuracy are significant, we follow [Clark and Ravazzolo \(2013\)](#) and apply [Diebold and Mariano \(1995\)](#) t -tests for equality of the average loss (with loss defined as squared error or CRPS). In the tables and figures, differences in accuracy that are statistically different from zero are denoted by one or two asterisks, corresponding to significance levels of 10% and 5%, respectively. The underlying p -values are based on t -statistics computed with a serial correlation-robust variance, using the pre-whitened quadratic spectral estimator of [Andrews and Monahan \(1992\)](#). Since our models are nested to the AR benchmark, we report p -values based on one-sided tests and look for rejection of the null of equal accuracy versus the alternative that the factor model is superior to the benchmark AR model.⁶

Finally, we evaluate the predictive densities using a test of absolute forecast accuracy. As in [Diebold et al. \(1998\)](#), we utilize the Probability Integral Transforms (PITS) of the realization of the variable with respect to the forecast densities. A forecast density is preferred if the density is correctly calibrated, regardless of the forecaster’s loss function. The PIT at time $\tau + 1$ are:

$$PIT_{k,\tau+1} = \int_{-\infty}^{y_{\tau+h}} p(\tilde{u}_{k,\tau+h}|y_{1:\tau}) d\tilde{u}_{k,\tau+h}. \quad (11)$$

and should be uniformly, independently (if $h = 1$) and identically distributed if the forecast densities $p(\tilde{y}_{k,\tau+h}|y_{1:t})$, for $\tau = \underline{t}, \dots, \bar{t}$, are correctly calibrated. Hence, calibration evaluation requires the application of tests for goodness of fit. We apply the [Berkowitz \(2001\)](#) test for zero mean, unit variance and independence of the PITS. The null of the test is no calibration failure. [Mitchell and Wallis \(2010\)](#) discuss the value of information-based methods for evaluating forecast densities that are well calibrated on the basis of PIT tests.

3 Results

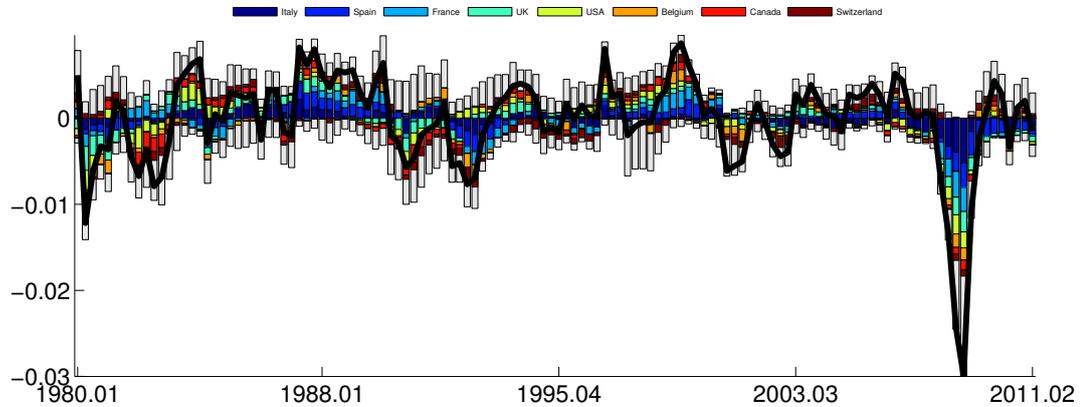
Below we first discuss the forecast performance of the global model for predicting GDP. In line with the in-sample business cycle synchronization literature, referred to in [Section 1](#), we focus on average results across all countries and on average results within geographical regions. We then expand the model to incorporate regional business cycle factors and evaluate the forecast performance of the expanded model. Finally, we incorporate factor uncertainty by combining factor model forecasts from models including up to 4 common business cycle factors.

3.1 Forecasting GDP using one global component

[Figure 2](#) reports the estimated global business cycle factor. The solid black line displays the factor estimate from the last forecast vintage, i.e., 2011:Q2. The white bars together with the coloured bars report the contribution to the factor estimate, at each

⁶The AR is estimated using Gibbs simulations, using the same priors as specified for the serially correlated idiosyncratic errors of the factor model, see [Sections 2.2](#) and [A.0.4](#).

Figure 2: The global business cycle factor and relative historical forecast performance GDP



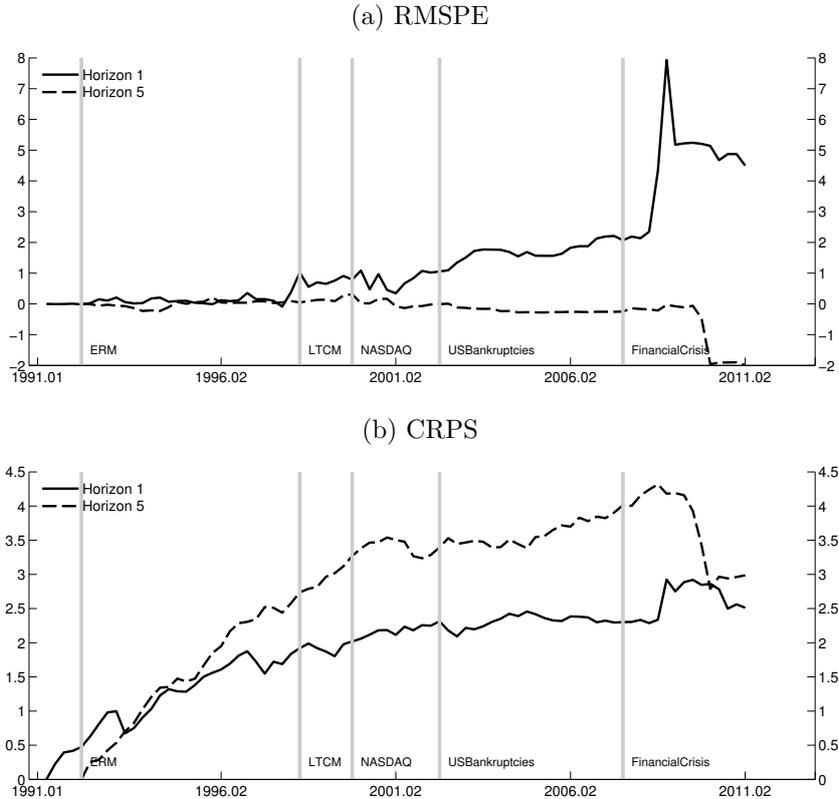
Note: The plot reports the global business cycle factor, estimated using the whole sample.

point in time, from the individual series used to derive the factor. In particular, the coloured bars show the eight most important series in terms of MSE over the entire sample. The contribution of each series is estimated based on the difference between the predicted state estimate and the updated state estimate within the Kalman Filter. All results are based on median estimates. As seen in the figure, the estimated factor has characteristics associated with the global business cycle showing a decline in world activity during the early 1980s and early 1990s, following the dot-com bubble that burst in 2000/2001, and during the Great Recession. The latter trough is by far the most severe. These are all periods that correspond closely to the recessions dated by the NBER for the US. The eight most important countries (in terms of MSE) are also primarily North American and European countries.

Figures 3a and 3b report the relative out-of-sample forecasting performance for GDP in all countries based on root mean square prediction error (RMSPE) and the continuous rank predictability score (CRPS), respectively, at forecast horizon 1 (solid line) and for horizon 5 (dotted line). In each plot the lines are the average relative forecast performance across all 33 countries in the sample. Particularly for RMSPE (CRPS) comparisons, the plots show the cumulative squared prediction errors of the benchmark; the AR(1), minus the cumulative squared prediction error of the Alternative model (i.e., the global factor model). Hence, an increase in a line indicates better performance of the global model; a decrease in a line indicates better performance of the AR(1). When the line is above zero the Alternative model has the best average forecast performance up to that point in time. In the plots we also report, with light grey bars, specific episodes that in the literature are considered as important events that might have significant effects on business cycles around the world.⁷

⁷These are: ERM; European Exchange Rate Mechanism crisis in 1992, LTCM; Long-Term Capital Management collapse in 1998-1999, NASDAQ; the Nasdaq crash in 1999-2000, US bankruptcies; wave of bankruptcies in the US in 2002, the Financial Crisis; the Global Recession spreading around the world in 2008, see [Imbs \(2010\)](#) for details.

Figure 3: Relative historical forecast performance GDP - one global component



Note: Figures 3a and 3b report relative out-of-sample forecasting performance, measured by the median across all countries. In particular, for RMSPE (CRPS) comparisons the plots show the cumulative squared prediction errors (the CRPS) of the benchmark minus the cumulative squared prediction error (the CRPS) of the Alternative model. An increase in a line indicates better performance of the Alternative model; a decrease in a line indicates better performance of the benchmark. The vertical grey lines identify some episodes with potential contagion effects.

Figure 3 has two main messages: The Great Recession was the first truly global recession in decades, and this time is different. In particular, Figure 3a shows that at horizon 1, a model including one global factor improves forecast performance (i.e., an increase in the line) relative to the benchmark. For RMSPE evaluation, the results are in particular strong when we include the period of the financial crisis, as seen by the sharp increase in the line early in the financial crisis. Although forecast performance increases steadily from the late 1990s (the Asian crisis), no other major business cycle event, e.g., the LTCM or the NASDAQ crash, caused such abrupt changes in relative forecast performance. Hence, the one-step ahead forecast performance is greatly improved using information contained in the global business cycle factor. Furthermore, the fact that the forecast performance of the global model increases sharply early in the financial crisis is consistent with the interpretation that the financial crisis is due to a common global shock, which may have affected most economies in a similar way. Hence, using a forecasting framework we confirm what Imbs (2010) has shown in an in-sample business cycle setting: The Great Recession was the first truly global recession in decades. We also confirm the in-sample evidence reported in Figure 1, namely

Table 1: Density calibration statistics: PIT test - one global component

	Horizon 1-step		5-step	
	Benchmark	Alternative	Benchmark	Alternative
All	0.15	0.30	0.12	0.33
North America	0.00	0.67	0.00	1.00
Asia	0.23	0.00	0.23	0.08
Europe	0.17	0.58	0.08	0.50
South America	0.00	0.20	0.00	0.20

Note: Each entry shows the fraction of density forecasts passing the Berkowitz (2001) test for zero mean, unit variance and independence of the PITs. The null of the test is no calibration failure, and we have used the 5 percent significance level as a cut-off in computing the fractions in the table. Evaluation sample: 1991:Q1-2011:Q2.

that business cycle synchronization increased during this period.⁸ It is also interesting to note that the short-term forecasting performance of the Alternative model relative to the benchmark declines again in the recovery phase of the financial crisis. This is consistent with the in-sample synchronization results reported in Figure 1, but is a new finding relative to that reported in Imbs (2010), where the sample ended in 2009:M5.

For density forecast evaluation, see Figure 3b, the results reported above hold for horizon 1. That is, the global factor model’s forecast performance relative to the benchmark increases almost monotonically already from the start of the evaluation sample. Yet, the increase in the performance of the global factor model relative to the AR(1) early in the financial crisis stands out.

Turning to the longer forecasting horizon, horizon 5, the forecast performance of the global model and the AR(1) are basically identical throughout the evaluation sample, based on RMSPE, or increase monotonically throughout the sample, based on CRPS. Thus, the Alternative model is better able than the benchmark to correctly capture the whole forecast distribution. If the forecast user is more concerned about higher order moments, this is important information and should, all else equal, favour the global factor model more relative to the simple AR(1).

A few years into the crisis, however, the five quarter ahead forecast performance deteriorates sharply. That is, from 2009/2010, the forecast performance of the global model declines sharply relative to the AR(1). This holds for both the RMSPE and the CRPS criteria. Hence, this time is different; Only after the Great Recession do we observe a sharp deterioration of the long horizon forecast performance of the global

⁸Engel and Wang (2011) has shown that trade in durable goods is an important element in open-economy rational expectations (RE) macro models that can account for some of the high correlation of output across countries. Yet, and as pointed out by the same authors, the channels explored may be different from those in the traditional RE models. One possibility is that agents receive strong signals about the future common component. If news helps to drive business cycles (c.f., Beaudry and Portier (2007)), then perhaps news about the common component also effectively filters an otherwise noisy signal, thereby increasing forecast performance across countries.

factor model relative to the benchmark. Why is this so? One explanation could be that while countries are synchronized in the recession phase of the financial crisis (when the shock occurs), the recovery phase has been much less synchronized. Some small open economies have recovered fast while others have recovered much more slowly, and some are still in a recession. To the extent that forecast errors are a good proxy for uncertainty, this is consistent with the findings in [Baker et al. \(2013\)](#), who find that since 2008/2009 uncertainty about the US fiscal situation and economic policies in Europe, in particular, has surged and slowed down the pace of an already slow recovery in these countries and areas. In any case, these idiosyncratic developments favour the AR(1) model relative to the model entertaining one common global component.

The first row of [Table 1](#) reports the results for the PIT tests, summarized as averages across all countries. The null hypothesis of correct calibration can not be rejected in only 15 (12) percent of the cases for the benchmark model on horizon 1 (horizon 5). For the Alternative model, i.e., the factor model, the performance is better, and rejection of the null hypothesis of correct calibration is obtained for 30 (33) percent of the cases on horizon 1 (horizon 5). Thus, the AR(1) model delivers less calibrated predictive densities on average. We notice that the variance of the predictive densities from the DFM is on average smaller than the one from the AR(1) model. The information content in the common global component reduces the forecast errors, shrinking the densities and increasing calibration. Therefore, global factors not only improve relative predictability, but also, and importantly, provide useful information that delivers higher calibration.

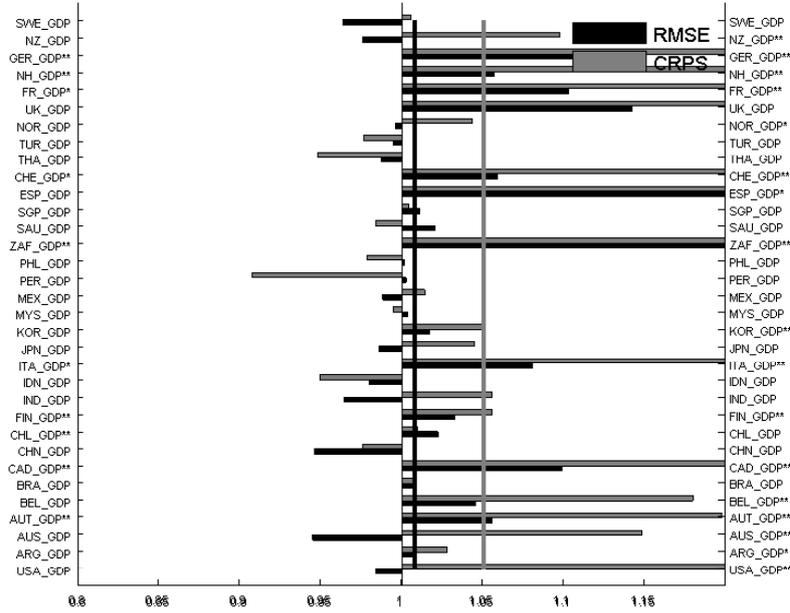
3.2 Country and region specific details

So far we have examined to what extent the global model can improve forecast performance for GDP relative to an AR(1) by looking at averages across all countries. Such aggregates can easily conceal interesting information. In this section we examine the country and region specific details behind [Figure 3](#), as well as tests of significance in terms of difference in forecasting performance. To organise the discussion for the individual countries, we focus on performance up until the financial crisis, and then on the period thereafter, see [Figures 4](#) and [5](#) for horizons 1 and 5, respectively. [Figures 8](#) and [9](#) decompose the results reported in [Figure 3](#) into regional averages. For brevity, the figures with regional details are reported in [Appendix B](#).

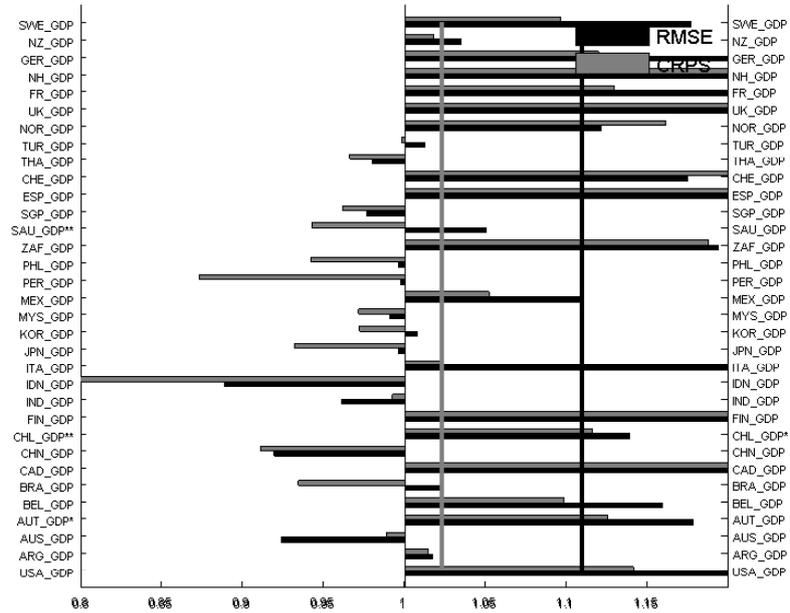
Country specific details confirm the picture from above. On average, the short-term forecasting performance of the global model increases before and immediately after the Great Recession. Focusing on the period up until 2007:Q3, we see that for many of the individual countries the performance of the Alternative model is also significantly better than the benchmark. This is especially so for CRPS evaluation. Among the countries where the global factor model does not seem to add much value in terms of short-term forecasting are Turkey, Thailand, Peru, Indonesia, and China, see [Figure 4a](#). Interestingly, these countries are also very different from the other countries in the sample in that they have exceptionally high and (or) volatile growth rates, see [Table 3](#) in [Appendix B](#). As the common factor captures commonalities across countries, it is not surprising that it does not add value in terms of forecasting performance for these countries.

Figure 4: Forecast evaluation GDP: 1 Factor, Horizon=1

(a) 1990:Q1-2007:Q3



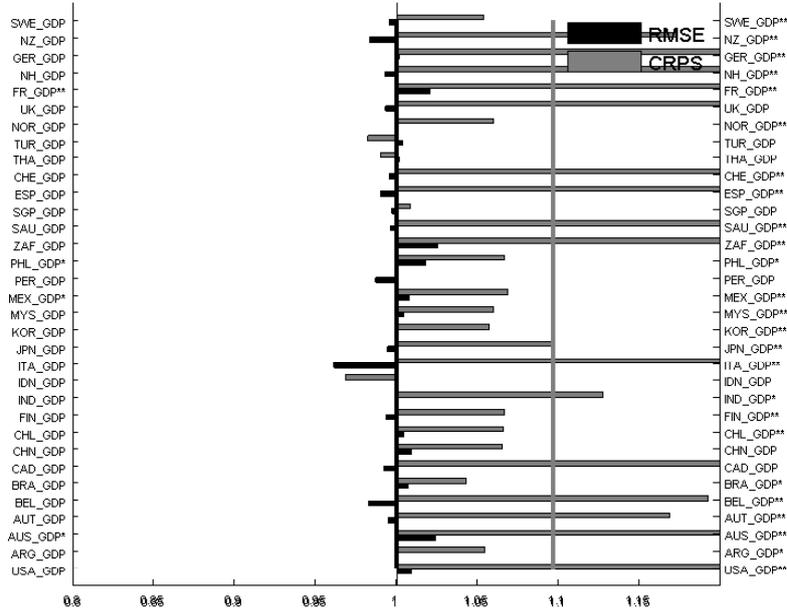
(b) 2007:Q4-2011:Q2



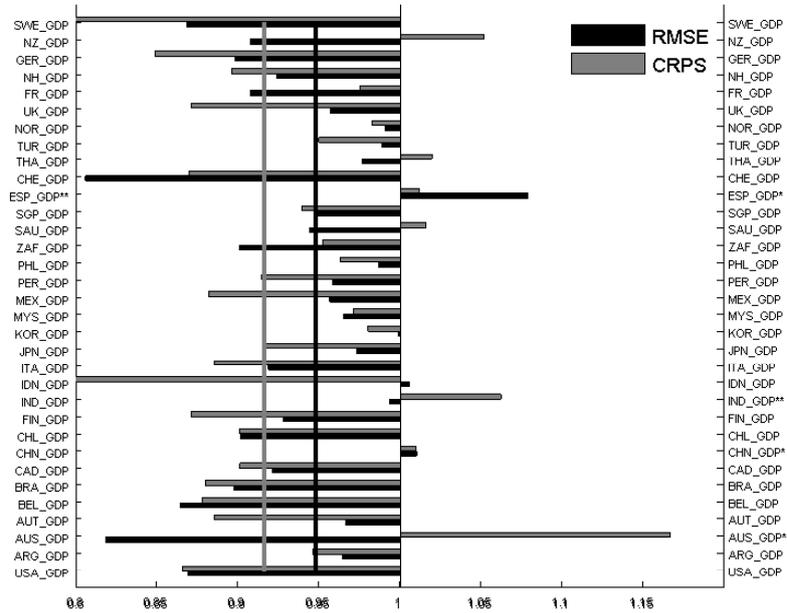
Note: The bars show the relative forecast performance of the Alternative models against the benchmark, normalized such that a value larger than 1 indicates that the Alternative model is better. The vertical lines report the average relative score across a given set of variables, as indicated by the line's coverage. The left- (right-) hand side y-axis reports the variable names together with Diebold and Mariano (1995) t-tests for equality of the RMSPE (CRPS). Based on one-sided tests a rejection of the null of equal accuracy versus the alternative that the Alternative model is superior to the benchmark is shown by one or two asterisks, corresponding to significance levels of 10% and 5%, respectively.

Figure 5: Forecast evaluation GDP: 1 Factor, Horizon=5

(a) 1990:Q1-2007:Q3



(b) 2007:Q4-2011:Q2



Note: See Figure 4.

Turning to longer-term forecasting, almost all countries benefit from entertaining the Alternative model, but as emphasized above, more so for density forecasting than for point forecasting, see Figure 5a. Going into the crisis, however, the picture is

reversed, and now almost all countries do worse when entertaining the global model relative to the benchmark.

Newer studies in the business cycle synchronization literature highlight a growing importance of regional business cycle developments relative to global one, see, e.g., [Crucini et al. \(2011\)](#), [Mumtaz et al. \(2011\)](#) and [Thorsrud \(2013\)](#). This motivates grouping the results reported above into regional averages and investigating to what extent the relative forecast performance for countries across regions differs.⁹ Doing so, we see from Figures 8 and 9, in Appendix B, that the Great Recession caused an unprecedented increase in the performance of the Alternative model relative to the benchmark at horizon 1, but a fall on longer forecasting horizons in all regions. This is in line with the more aggregated results reported above. However, the regional results also uncover differences in terms of the performance path across time and in absolute performance. In particular, for Asia, the global factor model is on average outperformed by the simple AR(1) throughout much of the evaluation sample. This should come as no surprise given that many outliers found in Figure 4a and 4b were from Asia. For the South American countries, on the other hand, there is an increase in the relative performance of the global model, but only until the late 1990s.

Large regional differences in relative forecasting performance can also be observed by looking at the PIT tests, see the second to fifth row of Table 1.¹⁰ Starting with Asia we see from the table that for almost all countries in this region, the density forecasts are badly calibrated, i.e., they do not pass the test. The benchmark model, the AR(1), does a better job with 23 percent of the countries passing the test (both horizons). This confirms the results reported above, emphasising the weak forecast performance when using the global model for many Asian countries. However, for the majority of countries in North America and Europe, we cannot reject the null hypothesis of no calibration failure. Moreover, the Alternative model outperforms the benchmark by a large margin. This also holds for countries in South America on average, but the fraction of countries in the region actually passing the test is only 20 percent.

3.3 Forecasting GDP including regional components

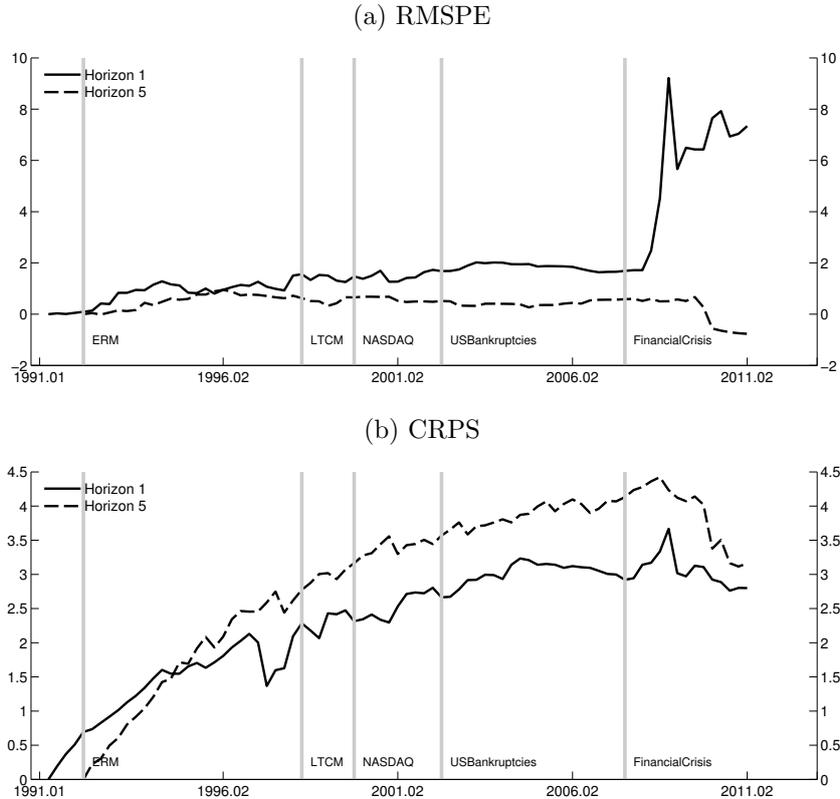
The results, and literature, reported above naturally beg the question: Can we improve the relative forecasting performance of the factor model by including region specific business cycle factors? To address this question we re-specify the global factor model such that it includes four regional business cycle factors: A North American factor, an Asian factor, a European factor, and a South American factor. The four factors are identified by employing the following ordering of the first four variables in y_t : the US, Korea, Germany, Brazil. See Section 2 for details about the identification strategy. Here we note that other alternative factor identification schemes could have been employed, see, e.g., [Bai and Wang \(2012\)](#). We prefer the one selected because it leaves the VAR(h) dynamics completely unrestricted and allows for spillovers between the regional factors. Accounting for such spillovers is consistent with findings in, e.g., [Thorsrud \(2013\)](#).

The four factors, as estimated using the whole sample, are reported in Figure 12

⁹It also motivates including regional business cycle factors into the model. We do so in the next section.

¹⁰Table 4, in Appendix B, reports the p-values for each individual country.

Figure 6: 4 factor model relative to AR(1)



Note: Figures 3a and 3b report relative out-of-sample forecasting performance, measured by the median across all countries. In particular, for RMSPE (CRPS) comparisons the plots show the cumulative squared prediction errors (the CRPS) of the benchmark minus the cumulative squared prediction error (the CRPS) of the Alternative model. An increase in a line indicates better performance of the Alternative model; a decrease in a line indicates better performance of the benchmark. The vertical grey lines identify some episodes with potential contagion effects.

in Appendix B. We note five silent facts: First, the North American factor resembles the one global factor used in Section 3.1. Second, the Great Recession is visible in all four factor estimates. Third, the Asian crisis around 1998 is clearly visible in the Asia-specific business cycle factor and not in any of the other estimates. Fourth, the European business cycle factor resembles the North American business cycle factor, questioning the presence of any truly common European specific business cycle factor, see, e.g., Canova et al. (2007). Finally, the South American business cycle shows large and volatile swings in the early part of the sample, consistent with the period when many of the South American countries were fighting hyper-inflation and particularly unstable macroeconomic developments.

The results of re-doing the out-of-sample forecasting experiment, but now with the factor model including four regional factors as the Alternative model, are reported in Figures 6a and 6b, for RMSPE and CRPS scores, respectively. As in Figure 3, we here only report averages across all countries. The effect of the Great Recession dominates both the short-run and long-run relative forecast performance. However, the absolute forecasting performance, relative to the AR(1), seems to become somewhat better

after including the additional factors, at least for RMSPE evaluation. That is, the distance from the zero line is larger (smaller, when negative) than in Figures 3a and 3b. Moreover, the performance does not seem to deteriorate as much at the end of the evaluation sample.

Detailed results for all the four regions are reported in Figures 10 and 11 in Appendix B. For the countries in North America and in Europe we confirm that the absolute forecasting performance, relative to the AR(1), has become somewhat better after including the additional factors. Also for South America, we observe that for RMSPE evaluation the absolute relative performance becomes better with the augmented factor model, both for short-run and long-run forecasting. However, looking at CRPS evaluation a different picture emerges. The performance of the augmented model relative to the AR(1) is worse than it was for the more parsimonious one-factor model, see Figure 11. Moreover, the effect of the Great Recession on relative forecasting performance is not visible. Finally, for Asia, we see that the relative short-run point forecasting performance of the augmented model clearly improves. Especially evident are the effects of the Asian crisis, resulting in a large and lasting increase in relative forecasting performance. The same dramatic shift can be seen when looking at Figure 11 and CRPS evaluation. Thus, including an Asia-specific business cycle factor improves forecasting performance for the Asian countries. However, on longer forecasting horizons, and for both point and density forecast evaluation, the augmented model actually seems to perform worse than the one-factor model relative to the AR(1).

The contrasting results of the augmented factor model for Asia and South American countries are surprising. One reason for these results might be that the more factor-rich model provides a better in-sample fit than the more parsimonious one-factor model, but that this fit, for the countries in Asia in particular, translates into worse relative forecasting performance at longer horizons. By introducing more factors into the model we also introduce more estimation uncertainty.

Table 2 reports the PIT tests for the factor model with four factors and the AR(1). As before, these models are denoted Alternative and Benchmark, respectively, in the table. Comparing the results in Table 2 to those in Table 1, we see that augmenting the factor model with four regional factors clearly improves the calibration of the density forecast, although not uniformly. On average across all countries 42 (36) percent now pass the test on horizon 1 (horizon 5). Most of the gain comes through a better calibration obtained for countries in Asia in particular, but also to some extent for countries in South America. Thus, our experiment with the augmented factor model delivers conflicting results: On the one hand, the relative forecasting performance seems to improve and also yield better calibrated densities for most countries. On the other hand, the inclusion of extra regional factors also seem to introduce extra noise in the model and forecasts. For some Asian and South American countries this reduces the gain obtained from using the augmented model.

In summary, the results reported in Sections 3.1 - 3.3 emphasize three regularities: First, in line with in-sample evidence showing an increase in cross-country business cycle synchronization during the Great Recession, we find a large and positive increase in the short-term relative forecasting performance of the global factor model during this period. However, our results indicate that this time is different; during the recovery phase of the recession relative forecasting performance declines, especially at longer forecasting horizons. Second, while the factor model on average outperforms the simple

Table 2: Density calibration statistics: PIT test - four regional components

	Horizon 1-step		5-step	
	Benchmark	Alternative	Benchmark	Alternative
All	0.15	0.42	0.12	0.36
North America	0.00	0.67	0.00	0.67
Asia	0.23	0.15	0.23	0.15
Europe	0.17	0.67	0.08	0.58
South America	0.00	0.40	0.00	0.20

Note: See the note to Table 1

benchmark model using both RMSPE and CRPS scoring rules, the factor model seems to be particularly good (relative to the benchmark) at correctly capturing the whole forecast distribution, i.e., when using the CRPS scoring rule. Third, regional factors matter, at least for short-term forecasting. Augmenting the one-factor model to include up to four regional business cycle factors yields an improvement in relative scores and gives better calibrated density forecasts. This result is consistent with business cycle synchronization studies documenting an increase in the importance of regional factors (relative to one global business cycle factor), but has not before been shown to hold in an out-of-sample forecasting experiment. Still, the latter results does not apply uniformly, suggesting that regional factors might play a varying role across countries and regions.

Our results offer an important extension to the existing in-sample business cycle synchronization literature. The results should also be of interest to model builders, policy makers and forecasters searching for which variables to include in their forecasting framework: Incorporating common global and regional factors increases forecasting accuracy.¹¹

3.4 Forecasting GDP using global components: Incorporating model uncertainty

The number of international business cycle factors to include in the model is uncertain, but matters for forecasting performance, as reflected by the results reported above. In this section we incorporate model uncertainty into the analysis by employing an out-of-sample model combination scheme. In line with the nature of the forecasting experiment we construct model weights based on the different factor model’s forecasting performance, and construct a combined factor model forecast based on these weights. The details are described below.

First, four factor models are estimated, each differentiated by the number of factors they include: One to four factors. The model with only one factor is identical to

¹¹More so for some countries than others. Moreover, as noted in Section 1, in a pure forecasting horse-race, other benchmark models (than the AR(1)) would very likely be harder to outperform. We leave it to future research to assess this for specific countries.

the model employed in Section 3.1. The model with two factors includes one North American and one Asia-specific business cycle factor. The model with three and four factors augments this model with a European and a South American business cycle factor, respectively.

We combine the different factor model forecasts using the linear opinion pool:

$$p(y_{\tau,h}) = \sum_{k=1}^K w_{k,\tau+h} g(y_{\tau+h}|I_{k,\tau}), \quad \tau = \underline{t}, \dots, \bar{t} \quad (12)$$

where K denotes the number of models to combine, $I_{k,\tau}$ is the information set used by model k to produce the density forecast $g(y_{\tau+h}|I_{k,\tau})$ for variable y at forecasting horizon h . \underline{t} and \bar{t} are the period over which the individual forecasters' densities are evaluated, and finally $w_{k,\tau+h}$ is a set of non-negative weights that sum to unity.

Combining the K density forecasts according to equation (12) can potentially produce a combined density forecast with characteristics quite different from those of the individual forecasters. As [Hall and Mitchell \(2007\)](#) notes, if all the individual forecasters' densities are normal, but with different mean and variance, the combined density forecast using the linear opinion pool will be mixture normal. This distribution can accommodate both skewness and kurtosis and be multimodal, see [Kascha and Ravazzolo \(2010\)](#).

We follow [Bjørnland et al. \(2011\)](#) and construct model weights according to:

$$w_{k,\tau+h} = \frac{h(s_{k,\tau})}{\sum_{i=1}^K h(s_{i,\tau})} \quad (13)$$

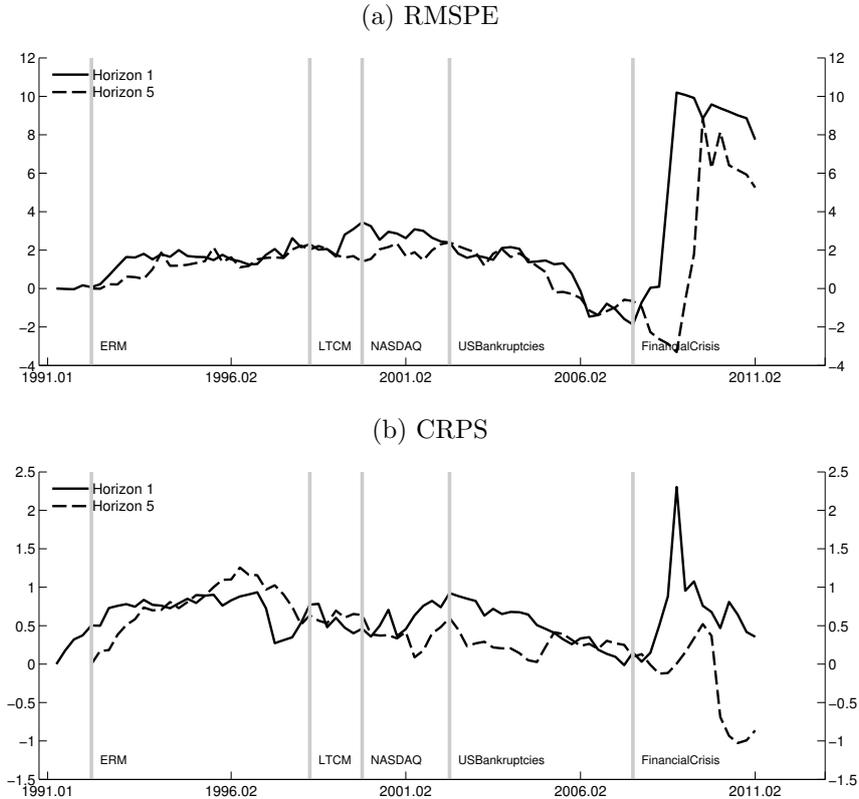
where $s_{i,\tau}$ is a statistic from the i th model at time τ , and $h(\cdot)$ is a monotonically increasing function. Two statistics are considered: the MSE and the CRPS. For both statistics the function $h(s) = \frac{1}{s}$. For MSE weights, s is computed using the square of equation (7). For CRPS weights s is computed using equation (9). Two points are worth emphasizing: The weights are derived based on out-of-sample performance, and the weights are horizon-specific.

Weighting schemes based on MSE weights are common in the model combination literature focusing on point forecasts, and has a long history, see, e.g., [Bates and Granger \(1969\)](#), [Clemen \(1989\)](#), and [Stock and Watson \(2004\)](#). Combining density forecasts has only more recently become popular, see, e.g., [Amisano and Giacomini \(2007\)](#), [Amisano and Geweke \(2009\)](#), [Kascha and Ravazzolo \(2010\)](#) and [Aastveit et al. \(2014\)](#) for recent applications. Still, using CRPS weights has so far not been common in the economic literature.¹²

To implement the model combination scheme, and to assess the effect of uncertainty related to how many factors to include in the model, we re-do the out-of-sample forecasting experiment described in Section 2.3. For each new vintage of data, we compute model weights as described by equation (13) and use these weights out-of-sample to construct a combined factor model forecast according to (12). Thus, the

¹²A commonly used weighting scheme in the density combination literature is based on the Logarithmic Score (LS). We prefer to use the CRPS measure mainly due to the fact that the CRPS is less sensitive to outliers compared to LS scoring, and because the CRPS also rewards values from the predictive density that are close but not equal to the actual realizations.

Figure 7: Combination relative to AR(1)



Note: The different plots report the average relative performance among all 33 countries. See also the notes to Figure 3.

weights will vary through time. The weights will also be horizon-specific. As before, the combined forecast is evaluated against an AR(1) as the benchmark.¹³

Addressing model uncertainty by employing model combination is often found to be preferable in empirical applications.¹⁴ Yet, model weights will be affected by estimation uncertainty. This is also seen in Figures 13 and 14, in Appendix B, which report the weights for horizons 1 and 5, respectively. Irrespective of how we construct the weights, the model weights are volatile and uncertain at the beginning of the evaluation sample, reflecting the limited information set on which they are derived.¹⁵ However, as more information is accumulated, the model weights seem to converge. In terms of constructing weights using MSE scores, we see from Figure 13 and for horizon 1 that the model including up to 4 business cycle factors gets a slightly higher weight than the other factor models, and that all the factor models also get a substantially

¹³The combined forecast is evaluated with RMSPE (CRPS) when the combination is obtained using MSE (CRPS) based weights. We stress that the weights are used out-of-sample. At the beginning of the evaluation sample, when no scores are available for weight computation, we simply use equal weighting.

¹⁴See, e.g., Timmermann (2006) for theoretical results and a discussion on when and how weighting will be optimal, and for more details about the motivation for doing model combination.

¹⁵Note that the weights reported are averages across all countries. As such, they under-report the actual variability observed in model weights across countries.

higher weight than the AR(1) model, consistent with the results already reported. On the longer forecasting horizon, horizon 5, it is almost impossible to distinguish the model weights from each other. Turning to CRPS based weights, Figure 14 shows that on both forecasting horizons the factor models are preferred to the simple AR(1) model. However, based on the model weights it is not easy to discriminate between the different factor models.

Figure 7 shows the relative performance of the combined factor model forecast. That is, the combined forecasts obtained using weights where the weight assigned to the AR(1) are normalized away. In terms of RMSPE evaluation, the performance of the combined factor model forecast outperforms the benchmark when evaluated over the whole sample. Compared to the results reported in Figure 3a, we also see that the combined forecast offers an improvement relative to the one-factor model (i.e., the distance from the zero line is larger (smaller, when negative)), on both forecasting horizons. Having noted that, the time path of the relative forecasting performance differs markedly from the one-factor model case. In particular, the precision of the long-run combined factor model forecast improves more substantially early in the crisis, and then remains more or less elevated. We do not observe any substantial drop in the short-term relative forecasting performance during the recovery phase of the Great Recession either. Thus, by combining the predictive content across all four factor models it seems that we are better able to capture country-specific (or region specific) information than in the one factor case. This resonates well with the interpretation where increased uncertainty about policies in the US and Europe are accompanied by large long-run forecast errors in the one factor model case (discussed in Section 3.1), but where the inclusion of additional regional factors (through model combination) reduces this uncertainty.¹⁶

Turning to the CRPS evaluation the main message from Section 3.1 holds through: The relative forecasting performance of the combined factor model forecast increases substantially on shorter horizons after the Great Recession, but falls some years into the crisis for longer horizons. However, compared to the results in Figure 3b, we now observe a much smaller gain in terms of CRPS scoring. Evaluated over the whole sample, the combined factor model forecast is actually outperformed by the AR(1) on horizon 5. As discussed in Section 3.3, one likely explanation for this deterioration of relative forecasting performance is the extra estimation uncertainty introduced by including additional factors. In the combination experiment, uncertainty related to the model weights adds to this estimation uncertainty. The combination results reported in Figure 7 indicate that these extra layers of model complexity are more harmful for density evaluation than for point forecast evaluation.

In sum, employing model combination to account for the uncertainty in the number of factors, offers no free lunch. Still, on average, regional business cycle factors matter and are given a large weight in the model combination experiment. The Great Recession has considerable impact on relative forecasting performance and highlights the potential benefit of using common international business cycle factors in forecasting.

¹⁶Remember here that in the one factor model case, the factor is identified using US GDP and that Figure 2 confirms that this factor is highly associated with European and North American countries.

4 Conclusion

The in-sample evidence pointing toward a high degree of co-movement in aggregate GDP across a large number of countries is well documented in the business cycle literature. However, high co-movement in-sample does not necessarily imply good out-of-sample performance, and inference without out-of-sample verification is likely to be spurious, with an out-of-sample approach inherently involving less over-fitting. In light of findings in, e.g., [Imbs \(2010\)](#), that argue that the Great Recession was the first really global recession in decades, understanding the nature of predictability using common global components seems especially relevant.

This paper therefore asks: Does a model with common international business cycle factors forecast better than the purely domestic alternative, i.e., a simple autoregressive process? To answer this question we employ a Dynamic Factor Model, commonly used in the business cycle synchronization literature, and run an out-of-sample forecasting experiment. We forecast GDP growth for a total of 33 countries and evaluate the forecast performance across 82 out-of-sample periods using both point and density evaluation measures.

Our results show that exploiting the informational content in a common global business cycle factor improves forecasting accuracy in terms of both point and density forecast evaluation across a large panel of countries. In line with in-sample evidence, we also document that the Great Recession has a huge impact on this result. Irrespective of which loss function we use, the event causes a clear preference shift towards the model including a common global factor. Similar shifts are not observed earlier in the evaluation sample. However, this time is different also in other respects. On longer forecasting horizons the performance of the DFM deteriorates substantially in the aftermath of the Great Recession. This indicates that the recession shock itself was felt globally, but that the recovery phase has been very different across countries. Finally, augmenting the DFM with regional factors improves the performance of the model further, giving out-of-sample support to the in-sample studies advocating the importance of regional business cycle factors. Still, when taking into account the uncertainty associated with which and how many regional factors to include in the model, the results are less clear cut: No factor model specification gets a substantially higher weight than the others, although the combined factor model forecasts seem to score slightly better than those obtained from a one factor model only.

References

- Aastveit, K. A., K. R. Gerdrup, A. S. Jore, and L. A. Thorsrud (2014). Nowcasting GDP in real time: A density combination approach. *Journal of Business & Economic Statistics* 32(1), 48–68.
- Alessi, L., E. Ghysels, L. Onorante, R. Peach, and S. Potter (2014). Central bank macroeconomic forecasting during the global financial crisis: the european central bank and federal reserve bank of new york experiences. *Journal of Business & Economic Statistics* 32(4), 483–500.
- Ambler, S., E. Cardia, and C. Zimmermann (2002). International transmission of the business cycle in a multi-sector model. *European Economic Review* 46(2), 273 – 300.
- Amisano, G. and J. Geweke (2009). Optimal prediction pools. Working Paper Series 1017, European Central Bank.
- Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. 25(2), 177–190.
- Andrews, D. and J. . Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.
- Ashley, R., C. W. J. Granger, and R. Schmalensee (1980). Advertising and aggregate consumption: An analysis of causality. *Econometrica* 48(5), 1149–1167.
- Backus, D., P. J. Kehoe, and F. E. Kydland (1995). International business cycles: Theory and evidence. In C. Plosser (Ed.), *Frontiers of Business Cycle Research*, pp. 331–257. Princeton University Press.
- Backus, D. K. and P. J. Kehoe (1992). International evidence of the historical properties of business cycles. *American Economic Review* 82(4), 864–88.
- Bai, J. and S. Ng (2010). Principal components estimation and identification of the factors. Unpublished manuscript.
- Bai, J. and P. Wang (2012). Identification and estimation of dynamic factor models. MPRA Paper 38434, University Library of Munich, Germany.
- Baker, S. R., N. Bloom, and S. J. Davis (2013). Measuring economic policy uncertainty. Chicago Booth research paper 13-02.
- Bates, J. and C. Granger (1969). The combination of forecasts. *Operations Research Quarterly* 20(4), 451–468.
- Baxter, M. and M. A. Kouparitsas (2005). Determinants of business cycle comovement: A robust analysis. *Journal of Monetary Economics* 52(1), 113–157.
- Beaudry, P. and F. Portier (2007). When can changes in expectations cause business cycle fluctuations in neo-classical settings? *Journal of Economic Theory* 135(1), 458–477.

- Berkowitz, J. (2001). Testing Density Forecasts, with Applications to Risk Management. *Journal of Business & Economic Statistics* 19(4), 465–74.
- Bjørnland, H. C., K. Gerdrup, A. S. Jore, C. Smith, and L. A. Thorsrud (2011). Weights and pools for a Norwegian density combination. *The North American Journal of Economics and Finance* 22(1), 61–76.
- Canova, F., M. Ciccarelli, and E. Ortega (2007). Similarities and convergence in G-7 cycles. *Journal of Monetary Economics* 54(3), 850–878.
- Carter, C. K. and R. Kohn (1994). On gibbs sampling for state space models. *Biometrika* 81(3), pp. 541–553.
- Clark, T. and F. Ravazzolo (2013). The macroeconomic forecasting performance of autoregressive models with alternative specifications of time-varying volatility. *Journal of Applied Econometrics* forthcoming.
- Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5.
- Crucini, M., A. Kose, and C. Otrok (2011). What are the driving forces of international business cycles? *Review of Economic Dynamics* 14(1), 156–175.
- D’Agostino, A., D. Giannone, and P. Surico (2006). (Un) Predictability and macroeconomic stability. Unpublished manuscript.
- Diebold, F. X., T. Gunther, and A. S. Tay (1998). Evaluating Density Forecasts with Applications to Finance and Management. *International Economic Review* 39, 863–883.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13(3), 253–63.
- Engel, C. and J. Wang (2011). International trade in durable goods: Understanding volatility, cyclical, and elasticities. *Journal of International Economics* 83(1), 37–52.
- Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102, 359–378.
- Gneiting, T. and R. Ranjan (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business and Economic Statistics* 29, 411–422.
- Groen, J., R. Paap, and F. Ravazzolo (2012). Real-time inflation forecasting in a changing world. *Journal of Business and Economic Statistics* 31(1), 29–44.
- Hall, S. G. and J. Mitchell (2007). Combining density forecasts. *International Journal of Forecasting* 23(1), 1–13.
- Imbs, J. (2010). The First Global Recession in Decades. *IMF Economic Review* 58(2), 327–354. Conference on Economic Linkages, Spillovers, and the Financial Crisis, Paris, France, January 28-29, 2010.

- Jurado, K., S. C. Ludvigson, and S. Ng (2013). Measuring uncertainty. NBER Working Papers 19456, National Bureau of Economic Research.
- Kascha, C. and F. Ravazzolo (2010). Combining inflation density forecasts. *Journal of Forecasting* 29(1-2), 231–250.
- Kim, C.-J. and C. R. Nelson (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications* (1 ed.), Volume 1. The MIT Press.
- Kose, M. A., C. Otrok, and C. H. Whiteman (2003). International business cycles: World, region, and country-specific factors. *American Economic Review* 93(4), 1216–1239.
- Mitchell, J. and K. F. Wallis (2010). Evaluating Density Forecasts: Forecast Combinations, Model Mixtures, Calibration and Sharpness. *Journal of Applied Econometrics* 26(6), 1023–1040.
- Mumtaz, H., S. Simonelli, and P. Surico (2011). International comovements, business cycle and inflation: A historical perspective. *Review of Economic Dynamics* 14(1), 176–198.
- Ravazzolo, F. and S. Vahey (2013). Forecast densities for economic aggregates from disaggregate ensembles. *Studies in Nonlinear Dynamics and Econometrics* forthcoming.
- Stock, J. H. and M. W. Watson (1999). Forecasting inflation. *Journal of Monetary Economics* 44(2), 293–335.
- Stock, J. H. and M. W. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* 20(2), 147–62.
- Stock, J. H. and M. W. Watson (2004). Combining forecasts of output growth in seven-country data set. *Journal of Forecasting* 23, 405–430.
- Stock, J. H. and M. W. Watson (2005). Understanding changes in international business cycle dynamics. *Journal of the European Economic Association* 3(5), 968–1006.
- Stockton, D. (2012). Review of the monetary policy committee’s forecasting capability. Report presented to the Court of the Bank of England, October.
- Thorsrud, L. A. (2013). Global and regional business cycles. shocks and propagations. (0012).
- Timmermann, A. (2006). *Forecast Combinations*, Volume 1 of *Handbook of Economic Forecasting*, Chapter 4, pp. 135–196. Elsevier.
- Welch, I. and A. Goyal (2008). A Comprehensive Look at the Empirical Performance of Equity Premium prediction. *Review of Financial Studies* 21(4), 253–303.

Appendices

Appendix A The Gibbs sampling approach

The three steps of the Gibbs sampler, described in Section 2.2, are iterated until convergence. Below we describe the three steps in more detail. The exposition follows [Kim and Nelson \(1999\)](#) (Chapter 8) closely, and we refer to their book for details.

For convenience, we repeat some notation: $\tilde{y}_T = [y_1, \dots, y_T]'$, $\tilde{f}_T = [f_1, \dots, f_T]'$, $H = [\lambda_0, \dots, \lambda_s]$, and $p_i = [\rho_{1,i}, \dots, \rho_{l,i}]$ for $i = 1, \dots, N$, and rewrite the state space model defined in equation 1 and 2 as:

$$y_t = \Lambda F_t + \epsilon_t \quad (14)$$

and

$$F_t = AF_{t-1} + e_t \quad (15)$$

where $F_t = [f'_t, \dots, f'_{t-h}]'$, $e_t = Gu_t$, with $u_t \sim i.i.d.N(0, Q)$ and:

$$A = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_h \\ I_q & 0 & \cdots & 0 \\ 0 & I_q & \ddots & \vdots \\ 0 & 0 & I_q & 0 \end{pmatrix}, \quad G = \begin{pmatrix} I_q \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \Lambda = (H \quad 0_{N, h-s}) \quad (16)$$

Note that $h > s$ in our application.

We also allow for serially correlated idiosyncratic errors. In particular, we consider the case where $\epsilon_{t,i}$, for $i = 1, \dots, N$, follows independent AR(1) processes:

$$\epsilon_{t,i} = p_i E_{t,i} + \omega_{t,i} \quad (17)$$

where $\omega_{t,i}$ is the AR(1) residuals with $\omega_{t,i} \sim i.i.d.N(0, \sigma_i^2)$,

$$R = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_N^2 \end{bmatrix}, \quad (18)$$

and $E_{t,i} = [\epsilon_{t-1,i}, \dots, \epsilon_{t-l,i}]'$.

A.0.1 Step 1: $\tilde{f}_T | \tilde{y}_T, \Lambda, A, R, Q, p$

We employ Carter and Kohn's multimove Gibbs sampling approach (see [Carter and Kohn \(1994\)](#)). Because the state space model given in equations 14 and 15 is linear and Gaussian, the distribution of F_T given \tilde{y}_T and that of F_t given F_{t+1} and \tilde{y}_t for $t = T-1, \dots, 1$ are also Gaussian:

$$F_T | \tilde{y}_T \sim N(F_{T|T}, P_{T|T}) \quad (19)$$

$$F_t | \tilde{y}_t, F_{t+1} \sim N(F_{t|t, F_{t+1}}, P_{t|t, F_{t+1}}), \quad t = T-1, T-2, \dots, 1 \quad (20)$$

where

$$F_{T|T} = E(F_T|\tilde{y}_T) \quad (21)$$

$$P_{T|T} = Cov(F_T|\tilde{y}_T) \quad (22)$$

$$F_{t|t,F_{t+1}} = E(F_t|\tilde{y}_t, F_{t+1}) = E(F_t|F_{t|t}, F_{t|t+1}) \quad (23)$$

$$P_{t|t,F_{t+1}} = Cov(F_t|\tilde{y}_t, F_{t+1}) = Cov(F_t|F_{t|t}, F_{t|t+1}) \quad (24)$$

Given $F_{0|0}$ and $P_{0|0}$, we obtain $F_{T|T}$ and $P_{T|T}$ from the last iteration of the Gaussian Kalman filter:

$$F_{t|t-1} = AF_{t-1|t-1} \quad (25)$$

$$P_{t|t-1} = AP_{t-1|t-1}A' + GQG' \quad (26)$$

$$K_t = P_{t|t-1}\Lambda'(\Lambda P_{t|t-1}\Lambda' + R)^{-1} \quad (27)$$

$$F_{t|t} = F_{t|t-1} + K_t(y_t - \Lambda F_{t|t-1}) \quad (28)$$

$$P_{t|t} = P_{t|t-1} - K_t\Lambda P_{t|t-1} \quad (29)$$

This means that at $t = T$ equation 28 and 29 above, together with equation 19, is used to draw $F_{T|T}$.

We draw $F_{t|t,F_{t+1}}$ for $t = T-1, T-2, \dots, 1$ based on 20, where $F_{t|t,F_{t+1}}$ and $P_{t|t,F_{t+1}}$ are generated from the following updating equations:

$$\begin{aligned} F_{t|t,F_{t+1}} &= E(F_t|F_{t|t}, F_{t|t+1}) \\ &= F_{t|t} + P'_{t|t}A(AP_{t|t}A' + GQG')^{-1}(F_{t+1} - AF_{t|t}) \end{aligned} \quad (30)$$

$$\begin{aligned} P_{t|t,F_{t+1}} &= Cov(F_t|F_{t|t}, F_{t|t+1}) \\ &= P_{t|t} + P_{t|t}A'(AP_{t|t}A' + GQG')AP_{t|t} \end{aligned} \quad (31)$$

A.0.2 Step 2: $A, Q|\tilde{y}_T, \tilde{f}_T, \Lambda, R, p$

Conditional on \tilde{f}_T , equation 15 is independent of the rest of the model, and the distribution of A and Q are independent of the rest of the parameters of the model, as well as the data.

By abusing notation, we put the transition equation in SUR form and define:

$$y = X\beta + \epsilon \quad (32)$$

where $y = [f_1, \dots, f_T]'$, $X = [X_1, \dots, X_T]'$, $\epsilon = [\epsilon_1, \dots, \epsilon_T]'$ and $\beta = [\beta_1, \dots, \beta_q]'$, with $\beta_k = [\phi_{1,k}, \dots, \phi_{h,k}]$ for $k = 1, \dots, q$. Further,

$$X_t = \begin{pmatrix} x_{t,1} & 0 & \cdots & 0 \\ 0 & x_{t,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & x_{t,q} \end{pmatrix}$$

with $x_{t,k} = [f'_{t-1}, \dots, f'_{t-h}]$. Finally, $\epsilon \sim i.i.d.N(0, I_q \otimes Q)$.¹⁷

¹⁷With the transition equation specified in SUR form it becomes easy to adjust the VAR(h) model such that different regressors enter the q equations of the VAR(h).

To simulate β and Q , we employ the independent Normal-Whishart prior:

$$p(\beta, Q) = p(\beta)p(Q^{-1}) \quad (33)$$

where

$$p(\beta) = f_N(\beta|\underline{\beta}, \underline{V}_\beta) \quad (34)$$

$$p(Q^{-1}) = f_W(Q^{-1}|\underline{v}_Q, \underline{Q}^{-1}) \quad (35)$$

The conditional posterior of β is:

$$\beta|y, Q^{-1} \sim N(\bar{\beta}, \bar{V}_\beta)_{I[s(\beta)]} \quad (36)$$

with

$$\bar{V}_\beta = (\underline{V}_\beta^{-1} + \sum_{t=1}^T X_t' Q^{-1} X_t)^{-1} \quad (37)$$

and

$$\bar{\beta} = \bar{V}_\beta (\underline{V}_\beta^{-1} \underline{\beta} + \sum_{t=1}^T X_t' Q^{-1} y_t) \quad (38)$$

$I[s(\beta)]$ is an indicator function used to denote that the roots of β lie outside the unit circle.

The conditional posterior of Q^{-1} is:

$$Q^{-1}|y, \beta \sim W(\bar{v}_Q, \bar{Q}^{-1}) \quad (39)$$

with

$$\bar{v}_Q = \underline{v}_Q + T \quad (40)$$

and

$$\bar{Q} = \underline{Q} + \sum_{t=1}^T (y_t - X_t \beta)(y_t - X_t \beta)' \quad (41)$$

A.0.3 Step 3: $\Lambda, R, p|\tilde{y}_T, \tilde{f}_T, A, Q$

Conditional on \tilde{f}_T , and given our assumption of R being diagonal, equation 14 result in N independent regression models.

However, to take into account serially correlated idiosyncratic errors, and still employ standard Bayesian techniques, we need to transform equation 14 slightly.

Thus, for $i = 1, \dots, N$, conditional on p , and with $l = 1$, we can rewrite equation 14 as:

$$y_{t,i}^* = \Lambda_i F_t^* + \omega_{t,i} \quad (42)$$

with $y_{t,i}^* = y_{t,i} - p_{1,i} y_{t-1,i}$, and $F_t^* = F_t - p_{1,i} F_{t-1}$, and Λ_i being the i -th row of Λ .

From 42 we can then simulate the parameters Λ_i and $R_{i,i} = \sigma_i^2 = \frac{1}{h_i}$ using standard independent Normal-Gamma priors (for notational convenience we drop the subscript i from the expressions below):¹⁸

$$p(\Lambda, h) = p(\Lambda)p(h) \quad (43)$$

¹⁸Note that with $l = 0$, we could have simulated the parameters Λ_i and σ_i^2 without doing the transformation of variables described above.

where

$$p(\Lambda) = f_N(\Lambda|\underline{\Lambda}, \underline{V}_\Lambda) \quad (44)$$

$$p(h) = f_G(h|\underline{s}^{-2}, \underline{v}_h) \quad (45)$$

The conditional posterior of Λ is:

$$\Lambda|\tilde{y}, h, p \sim N(\bar{\Lambda}, \bar{V}_\Lambda) \quad (46)$$

with;

$$\bar{V}_\Lambda = (\underline{V}_\Lambda^{-1} + h \sum_{t=1}^T F_t^{*'} F_t^*)^{-1} \quad (47)$$

and

$$\bar{\Lambda} = \bar{V}_\Lambda (\underline{V}_\Lambda^{-1} \underline{\Lambda} + h \sum_{t=1}^T F_t^{*'} y_t^*) \quad (48)$$

The conditional posterior for h is:

$$h|\tilde{y}, \Lambda, p \sim G(\bar{v}_h, \bar{s}^{-2}) \quad (49)$$

with

$$\bar{v}_h = \underline{v}_h + T \quad (50)$$

and

$$\bar{s} = \frac{\sum_{t=1}^T (y_t^* - \Lambda F_t^*)' (y_t^* - \Lambda F_t^*) + \underline{v}_h \underline{s}^2}{\bar{v}_h} \quad (51)$$

Finally, conditional on Λ and h , the posterior of p depends upon its prior, which we assume is a multivariate Normal, i.e.:

$$p(p) = f_N(p|\underline{p}, \underline{V}_p) \quad (52)$$

Accordingly, the conditional posterior for p is:

$$p|\tilde{y}, \Lambda, h \sim N(\bar{p}, \bar{V}_p)_{I[s(p)]} \quad (53)$$

with

$$\bar{V}_p = (\underline{V}_p^{-1} + h \sum_{t=1}^T E_t' E_t)^{-1} \quad (54)$$

and

$$\bar{p} = \bar{V}_p (\underline{V}_p^{-1} \underline{p} + h \sum_{t=1}^T E_t' \epsilon_t) \quad (55)$$

A.0.4 Prior specifications and initial values

The Benchmark model is estimated using two-step parameter estimates (see Section 2.2) as priors. We label these estimates OLS. In particular, for equations 34 and 35 we set $\underline{\beta} = \beta^{OLS}$, $\underline{V}^\beta = V_\beta^{OLS} \times 3$, $\underline{Q} = Q^{OLS}$ and $\underline{v}_Q = 10$.

For equations 44, 45 and 52 we set $\underline{v}_h = 10$, $\underline{s}^2 = s^{2,OLS}$, $\underline{\Lambda} = [\lambda_0^{OLS} : 0_{N,h-s-1}]$ and $\underline{V}^\Lambda = [(I_s \times 3) \otimes V_{\lambda_0^{OLS}}]$. $\underline{p} = 0$, and $\underline{V}^p = 0.5$.

In sum, these priors are reasonable uninformative, but still proper. We have also experimented with other prior specifications, e.g. using Minnesota style prior for the transition equation parameters, and setting $\underline{\Lambda} = 0$. This yields similar results as those reported in the main text. However, the variables in our sample display very different unconditional volatilities. The prior specification should accommodate this feature.

The Gibbs sampler is initialized using parameter values derived from the two-step estimation procedure. Parameters not derived in the two-step estimation (i.e. p and $\lambda_1, \dots, \lambda_s$) are set to 0.

In this model, a subtle issue arises for the $t = 0$ observations (i.e. lags of the dynamic factors and the idiosyncratic errors at time $t = 1$). However, since we assume stationary errors in this model, the treatment of initial conditions is of less importance. Accordingly, we follow common practice and work with the likelihood based on data from $t = h + 1, \dots, T$.

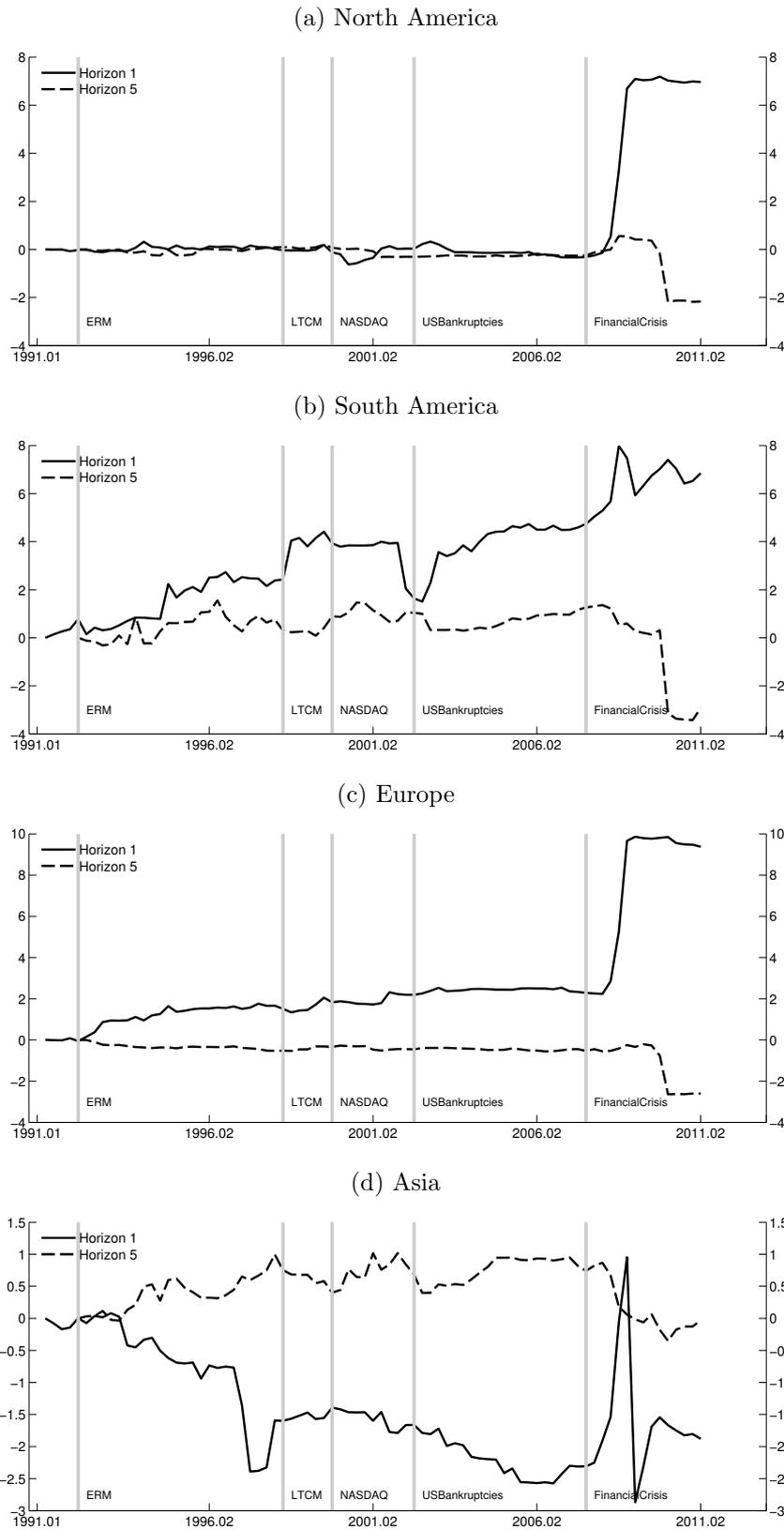
Appendix B Additional figures and tables

Table 3: Data and factor statistics

Country	Statistics Data			RMSE Factors			
	Mean	Std.	ACF	N. America	Asia	Europe	S. America
Argentina	0.64	2.19	1.09	0.58	1.01	0.64	0.71
Australia	0.79	0.75	0.99	0.56	1.03	0.65	0.69
Austria	0.54	1.13	0.64	0.57	1.03	0.63	0.72
Belgium	0.49	0.70	0.67	0.56	1.02	0.60	0.69
Brazil	0.70	1.77	0.05	0.57	1.03	0.64	0.70
Canada	0.63	0.76	0.42	0.50	1.03	0.65	0.66
Chile	1.13	2.00	0.48	0.57	1.02	0.64	0.69
China	2.47	1.14	0.35	0.57	1.03	0.65	0.72
Finland	0.59	1.39	0.41	0.56	1.03	0.65	0.71
France	0.44	0.50	0.23	0.55	1.00	0.59	0.72
Germany	0.42	0.88	-0.10	0.56	1.03	0.58	0.70
India	1.54	0.99	1.07	0.58	1.03	0.65	0.72
Indonesia	1.24	2.17	0.77	0.57	1.01	0.64	0.70
Italy	0.36	0.65	0.80	0.54	1.01	0.59	0.70
Japan	0.46	1.03	0.31	0.57	1.00	0.62	0.72
Korea	1.45	1.65	-0.03	0.56	0.82	0.62	0.71
Malaysia	1.42	1.62	0.23	0.56	0.87	0.61	0.65
Mexico	0.65	1.56	0.82	0.56	1.02	0.63	0.68
Netherlands	0.51	0.78	-0.03	0.56	1.02	0.62	0.70
New Zealand	0.53	0.94	1.16	0.56	1.01	0.65	0.69
Norway	0.68	1.21	0.06	0.57	1.03	0.64	0.71
Peru	0.78	3.09	0.05	0.57	1.03	0.65	0.72
Philippines	0.78	1.48	2.85	0.57	1.02	0.64	0.72
Saudi Arabia	0.53	2.13	0.14	0.57	1.02	0.64	0.71
Singapore	1.64	1.90	1.17	0.57	0.98	0.63	0.69
South Africa	0.61	0.84	0.57	0.56	1.02	0.64	0.66
Spain	0.58	0.55	-0.33	0.55	1.03	0.59	0.73
Sweden	0.54	1.32	1.84	0.56	1.03	0.64	0.71
Switzerland	0.44	0.68	0.98	0.56	1.01	0.62	0.68
Thailand	1.32	1.62	1.03	0.56	0.86	0.62	0.70
Turkey	1.02	2.68	-0.01	0.57	1.03	0.64	0.71
UK	0.47	0.70	0.26	0.50	1.00	0.64	0.72
USA	0.63	0.75	1.15	0.52	1.03	0.64	0.68
Median	0.63	1.14	0.48	0.56	1.02	0.64	0.70

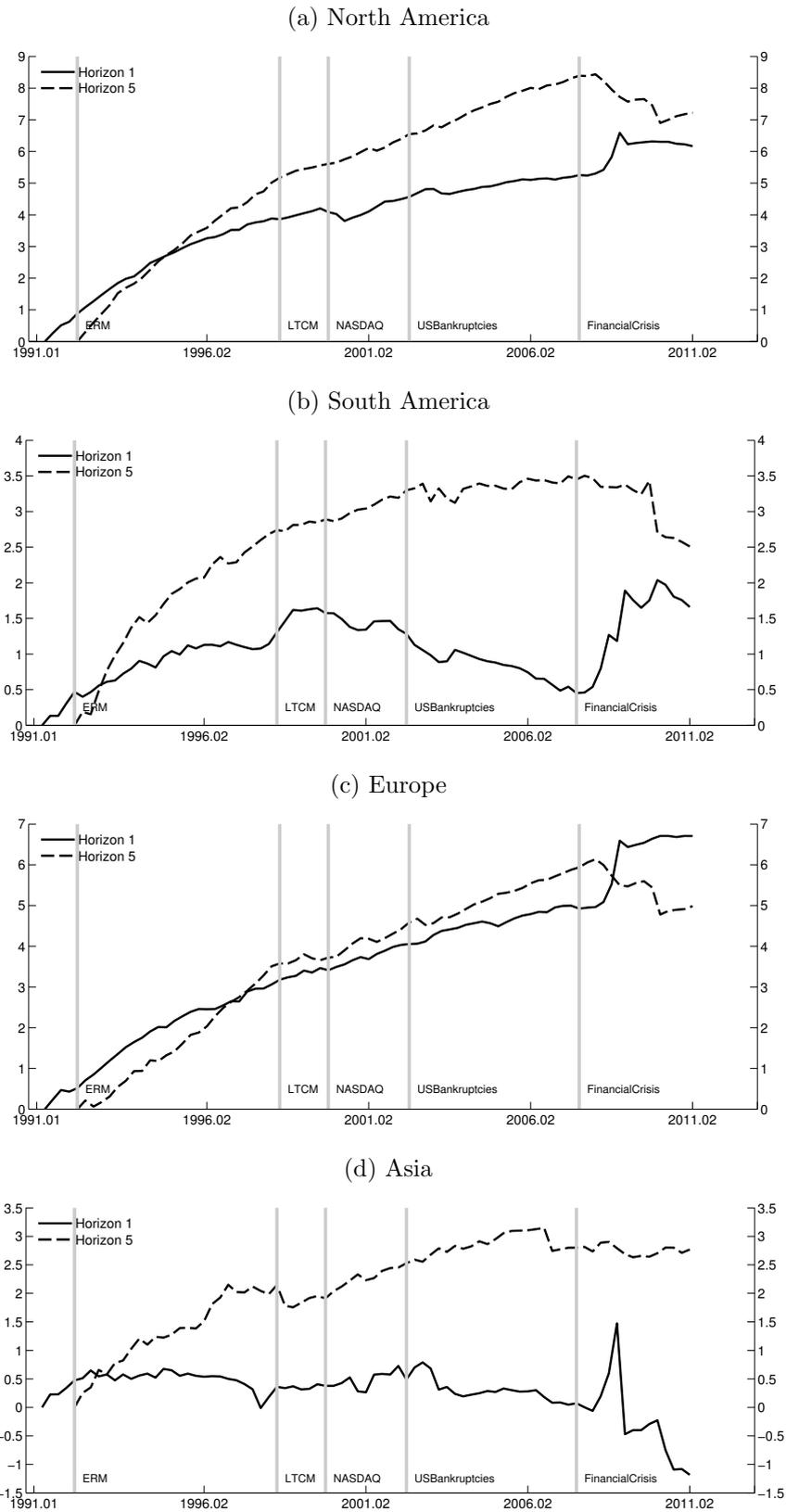
Note: Columns 2-4 report the mean, standard deviation (Std.) and autocorrelation (ACF) of the individual series. The ACF statistic is computed as the sum of the auto-correlation coefficients on lag 1-5. The columns associated with the RMSE Factors heading reflect the estimates based on estimating a BDFM with 4 regional factors. The RMSE columns report the root mean squares error associated with variable i in explaining factor j . In Figure 12 the coloured bars reflect the contribution by the 8 variables with the lowest RMSE.

Figure 8: Relative historical forecast performance GDP: Across regions, RMSPE scores



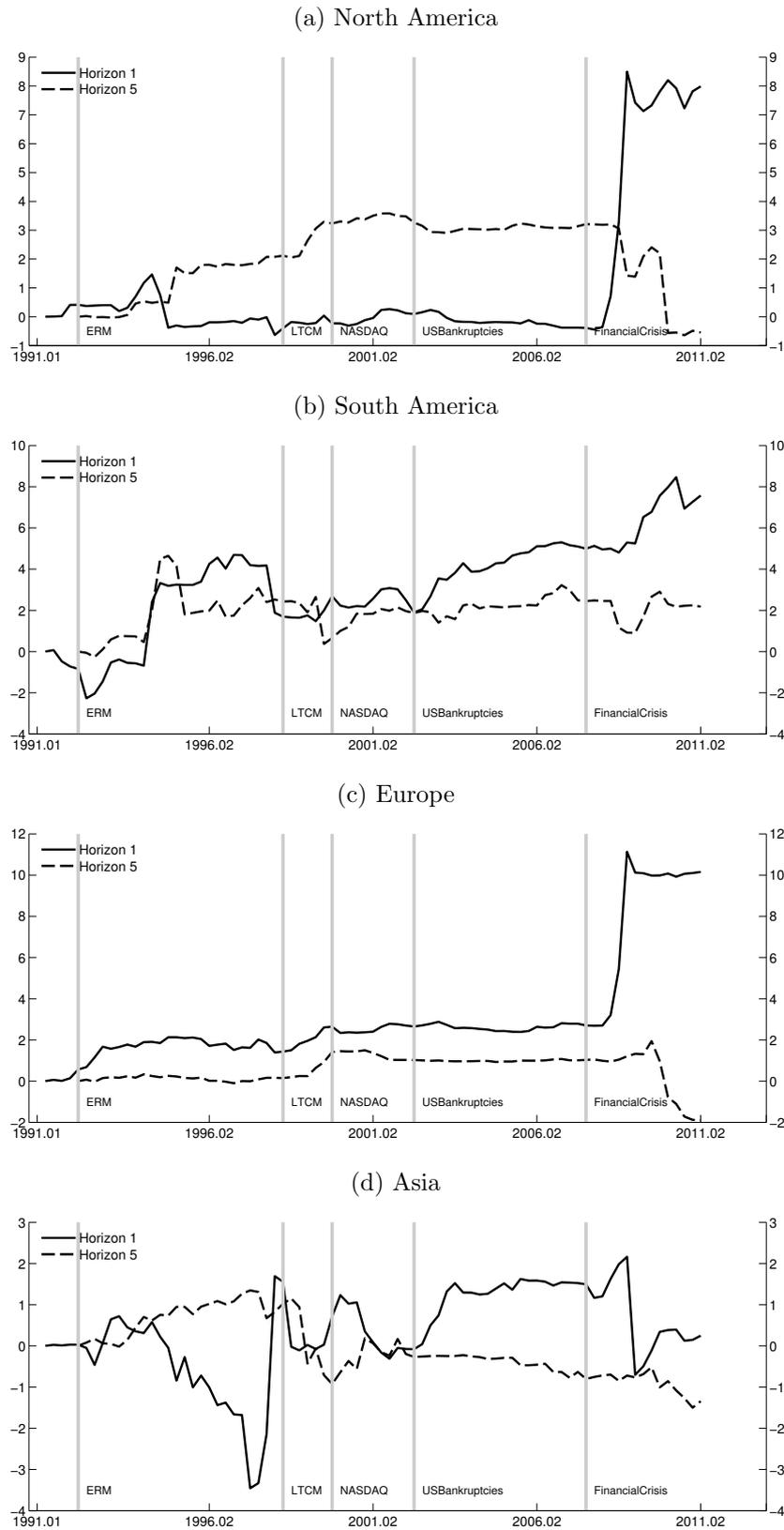
Note: The different plots report the average relative performance among countries within a geographical region. See also the notes to Figure 3.

Figure 9: Relative historical forecast performance GDP: Across regions, CRPS scores



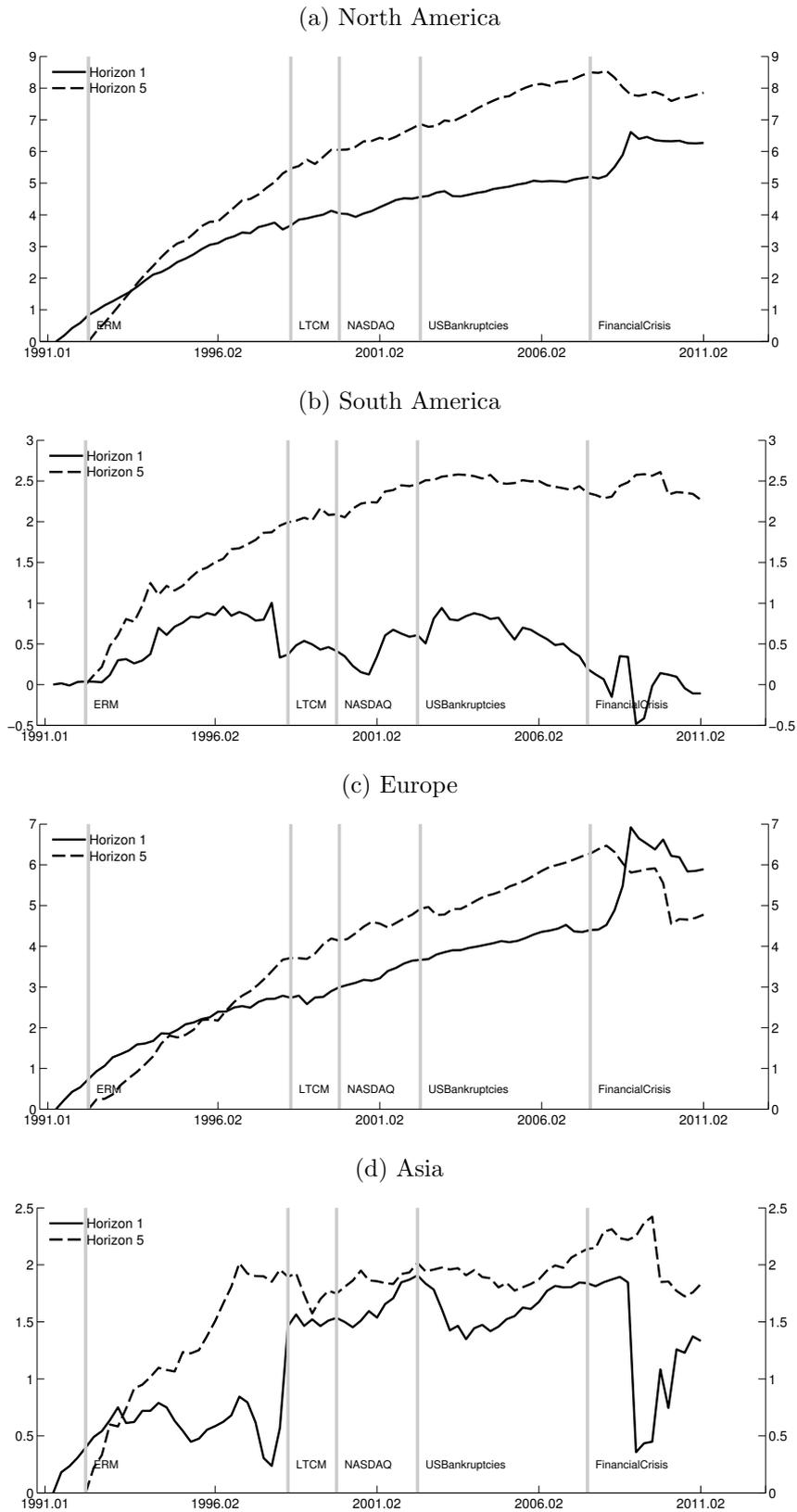
Note: The different plots report the average relative performance among countries within a geographical region. See also the notes to Figure 3.

Figure 10: 4 factor model relative to AR(1), RMSPE scores



Note: The different plots report the average relative performance among countries within a geographical region. See also the notes to Figure 3.

Figure 11: 4 factor model relative to AR(1), CRPS scores



Note: The different plots report the average relative performance among countries within a geographical region. See also the notes to Figure 3.

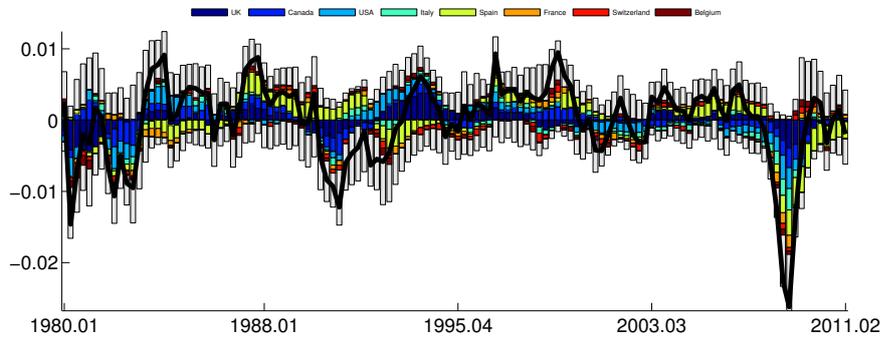
Table 4: Density calibration statistics: PIT test - one global component, all countries

	Horizon 1-step		5-step	
	Benchmark	Alternative	Benchmark	Alternative
USA	0.00	0.11	0.00	0.24
Argentina	0.02	0.38	0.00	0.00
Australia	0.00	0.00	0.00	0.01
Austria	0.67	0.86	0.03	0.00
Belgium	0.00	0.71	0.00	0.00
Brazil	0.00	0.02	0.00	0.17
Canada	0.00	0.00	0.00	0.59
China	0.02	0.00	0.00	0.00
Chile	0.00	0.00	0.00	0.00
Finland	0.14	0.66	0.23	0.31
India	0.00	0.00	0.00	0.01
Indonesia	0.00	0.00	0.08	0.52
Italy	0.00	0.00	0.00	0.00
Japan	0.00	0.00	0.00	0.00
Korea	0.00	0.02	0.00	0.02
Malaysia	0.47	0.01	0.01	0.00
Mexico	0.01	0.13	0.00	0.15
Peru	0.00	0.00	0.00	0.00
Philippines	0.00	0.00	0.00	0.00
South Africa	0.00	0.00	0.00	0.02
Saudi Arabia	0.00	0.00	0.00	0.00
Singapore	0.01	0.00	0.54	0.00
Spain	0.00	0.03	0.00	0.00
Switzerland	0.00	0.15	0.00	0.18
Thailand	0.06	0.00	0.00	0.00
Turkey	0.53	0.03	0.41	0.02
Norway	0.00	0.06	0.00	0.09
UK	0.00	0.00	0.00	0.06
France	0.00	0.55	0.00	0.00
Netherlands	0.00	0.00	0.00	0.64
Germany	0.00	0.54	0.00	0.32
New Zealand	0.00	0.00	0.00	0.00
Sweden	0.00	0.01	0.00	0.01

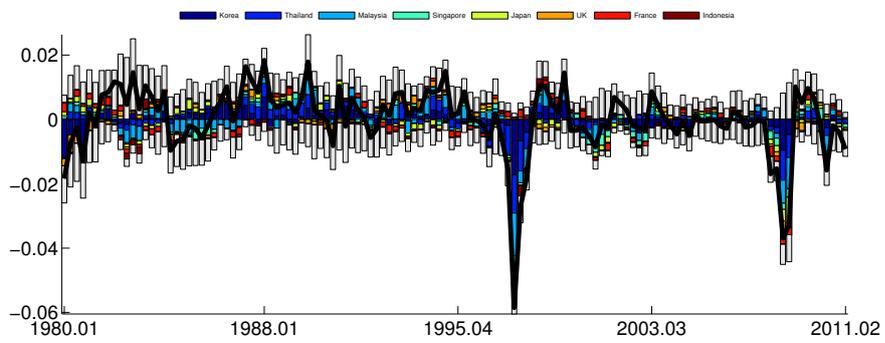
Note: Each entry shows the p -value of the [Berkowitz \(2001\)](#) test for zero mean, unit variance and independence of the PITS. The null of the test is no calibration failure. Evaluation sample: 1991:Q1-2011:Q2.

Figure 12: Factors, 4 factor model

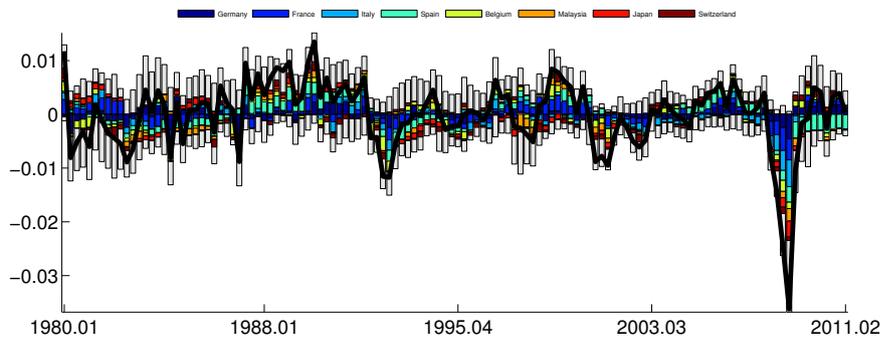
(a) 1



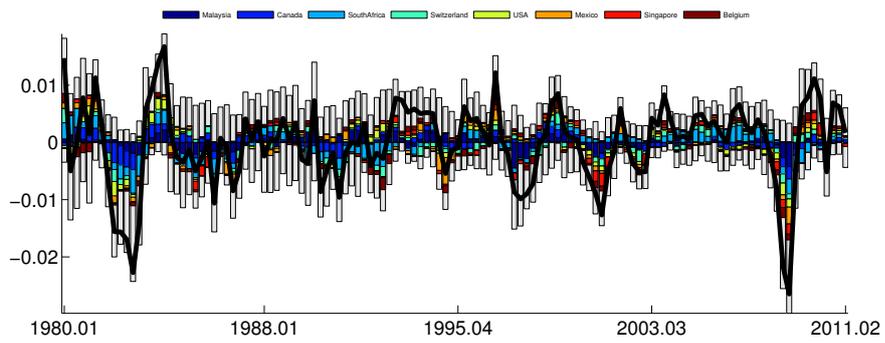
(b) 2



(c) 3

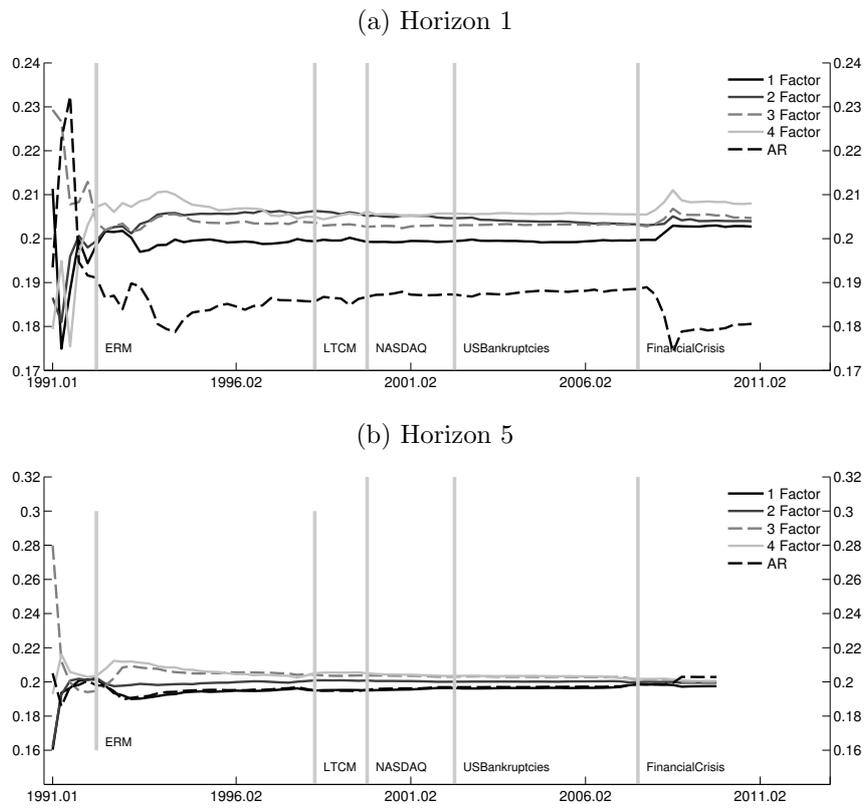


(d) 4



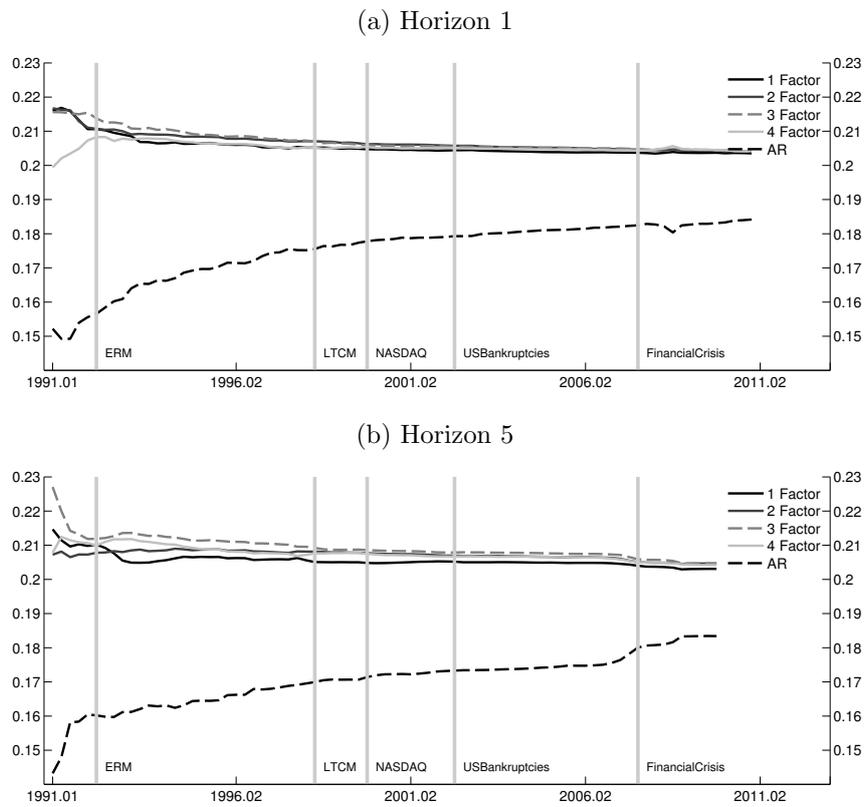
Note: The plots report 4 regional business cycle factors, estimated using the whole sample. See also Section 3.1.

Figure 13: Model weights: MSE



Note: Recursively estimated model weights.

Figure 14: Model weights: CRPS



Note: Recursively estimated model weights.