

How good are dynamic factor models at forecasting output and inflation? A meta-analytic approach

Sandra Eickmeier

(Deutsche Bundesbank and University of Cologne)

Christina Ziegler

(University of Bonn)



Discussion Paper
Series 1: Economic Studies
No 42/2006

Editorial Board:

Heinz Herrmann
Thilo Liebig
Karl-Heinz Tödter

Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-1

Telex within Germany 41227, telex from abroad 414431

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet <http://www.bundesbank.de>

Reproduction permitted only if source is stated.

ISBN 3-86558-235-4 (Printversion)

ISBN 3-86558-236-2 (Internetversion)

Abstract:

This paper surveys existing factor forecast applications for real economic activity and inflation by means of a meta-analysis and contributes to the current debate on the determinants of the forecast performance of large-scale dynamic factor models relative to other models. We find that, on average, factor forecasts are slightly better than other models' forecasts. In particular, factor models tend to outperform small-scale models, whereas they perform slightly worse than alternative methods which are also able to exploit large datasets. Our results further suggest that factor forecasts are better for US than for UK macroeconomic variables, and that they are better for US than for euro-area output; however, there are no significant differences between the relative factor forecast performance for US and euro-area inflation. There is also some evidence that factor models are better suited to predict output at shorter forecast horizons than at longer horizons. These findings all relate to the forecasting environment (which cannot be influenced by the forecasters). Among the variables capturing the forecasting design (which can, by contrast, be influenced by the forecasters), the size of the dataset from which factors are extracted seems to positively affect the relative factor forecast performance. There is some evidence that quarterly data lend themselves better to factor forecasts than monthly data. Rolling forecasts are preferable to recursive forecasts. The factor estimation technique seems to matter as well. Other potential determinants - namely whether forecasters rely on a balanced or an unbalanced panel, whether restrictions implied by the factor structure are imposed in the forecasting equation or not and whether an iterated or a direct multi-step forecast is made - are found to be rather irrelevant. Moreover, we find no evidence that pre-selecting the variables to be included in the panel from which factors are extracted helped to improve factor forecasts in the past.

JEL: C2, C3, E37

Key words: Factor models, forecasting, meta-analysis

Non-Technical Summary

In recent times, dynamic factor models are increasingly applied in central banks to forecast economic developments based on large datasets. This paper adopts a meta-analytic approach to evaluate the forecast performance of these models relative to other models for real economic activity and inflation. This approach allows us to systematically summarize findings from existing factor forecast applications using statistic and econometric techniques. At the same time, we are able to assess the relevance of a large number of determinants which are able to influence the forecast performance.

Based on more than 45,000 observations taken from 46 studies, we find that, on average, factor forecasts perform slightly better than other models' forecasts, with inflation forecasts having some advantages over output forecasts. However, this also involves a greater dispersion of results associated with inflation.

The main findings regarding the determinants are the following. The size of the dataset from which factors are extracted seems to positively affect the relative factor forecast performance. Recursive factor forecasts are inferior to rolling factor forecasts, which are able to better account for possible structural breaks in the factors. Moreover, it seems to matter which factor estimation technique is applied. According to our analysis, the dynamic approaches proposed by Forni et al. (2005) and Kapetanios and Marcellino (2004) tend to outperform the static approach of Stock and Watson (2002), which, however, is much easier to implement. Our results further suggest that factor models have some advantages when predicting US compared to UK macroeconomic variables. Factor models also seem to be better suited to forecasting US output than euro-area output, but there are no significant differences between the relative factor forecast performance for US and euro-area inflation. Factor models perform slightly worse than alternative methods which are also able to exploit large datasets, but they generally outperform small-scale models. There is some evidence that quarterly data are better suited to factor forecasts than monthly data and that the relative factor forecast performance worsens with the forecast horizon for output, whereas the forecast horizon does not seem to matter for inflation. Other potential determinants, namely whether forecasters rely on a balanced or an unbalanced panel, whether restrictions implied by the factor structure are imposed in the forecasting equation or whether this equation is estimated unrestrictedly and whether an iterated or a direct multi-step forecast is made, are found to be rather irrelevant. Moreover, we find no evidence that pre-selecting the variables to be included in the panel from which factors are extracted helps to improve factor forecasts.

Our analysis can help practitioners to improve their forecasts. Our main messages are the following. First, it seems worthwhile exploiting large datasets. Second, it is important to carefully specify the forecasting model. Although more demanding in terms of specification, complex factor estimation techniques tend to perform relatively well and restricted estimation of the forecasting equation does not seem to be inferior to unrestricted estimation. Third,

some potential determinants were found to be surprisingly irrelevant in the empirical analysis, given that researchers have paid relatively much attention to them in the past. This refers, for instance, to whether the data have been selected prior to the factor estimation, whether or not restrictions implied by the factor structure are imposed in the forecasting equation and whether an unbalanced or a balanced panel is used. Another determinant, by contrast, turns out to be surprisingly important and robust, given that it has not attracted much attention in the existing literature to date: whether the forecast is made based on a rolling or a recursive window.

Nicht technische Zusammenfassung

Dynamische Faktormodelle werden in der letzter Zeit verstärkt von Zentralbanken angewandt, um volkswirtschaftliche Entwicklungen auf Basis umfangreicher Datensätze vorherzusagen. Der vorliegende Artikel untersucht mit Hilfe einer Meta-Analyse die Fähigkeit solcher Modelle zur Prognose von realwirtschaftlicher Aktivität und Inflation relativ zur Prognosefähigkeit alternativer Modelle. Dieser Ansatz erlaubt es, auf Basis statistischer und ökonometrischer Methoden einen systematischen Überblick über existierende Anwendungen zur Prognose mit Faktormodellen zu geben. Gleichzeitig kann die Bedeutung einer großen Anzahl möglicher Determinanten untersucht werden, welche die Prognosegüte beeinflussen können.

Auf Basis von über 45,000 Beobachtungen, die aus 46 Studien entnommen wurden, finden wir, dass im Durchschnitt Faktormodelle leicht besser prognostizieren als andere Modelle, mit gewissen Vorteilen bei der Inflationsprognose gegenüber der Prognose der realwirtschaftlichen Aktivität; allerdings streuen die Ergebnisse bei der Inflationsprognose auch stärker.

Die wichtigsten Ergebnisse bezüglich der Determinanten lassen sich wie folgt zusammenfassen. Die Größe des Datensatzes, auf dessen Basis die Faktoren geschätzt werden, scheint sich positiv auf die relative Prognosefähigkeit von Faktormodellen auszuwirken. Rekursive Faktorprognosen sind rollierenden Faktorprognosen, die mögliche Strukturbrüche in den Faktoren besser berücksichtigen können, unterlegen. Des Weiteren scheint es eine Rolle zu spielen, welche Technik zur Schätzung der Faktoren angewandt wird. So schneiden die dynamische Schätzmethode von Forni et al. (2005) und von Kapetanios und Marcellino (2004) in unserer Analyse besser ab als der von Stock und Watson (2002) vorgeschlagene statische Ansatz, welcher allerdings den Vorteil hat, dass er leichter zu implementieren ist. Unsere Ergebnisse legen zudem nahe, dass Faktormodelle Vorteile bei der Prognose US-amerikanischer makroökonomischer Variablen haben im Vergleich zur Prognose britischer Variablen. Faktormodelle scheinen zudem besser geeignet, Output in den Vereinigten Staaten als Output im Euro-Raum vorherzusagen, wohingegen sich die Befunde zur Inflationsprognose in beiden Gebieten nicht signifikant voneinander unterscheiden. Wir finden weiter, dass mit Faktormodellen leicht schlechter prognostiziert wird als mit alternativen Methoden, welche ebenfalls in der Lage sind, große Datensätze zu nutzen; dagegen schneiden Faktormodelle in der Regel besser ab als kleine Prognosemodelle. Zudem gibt es Evidenz dafür, dass die Prognose mit Faktormodellen relativ zu anderen Modellen besser ist, wenn sie auf Basis vierteljährlicher als auf Basis monatlicher Daten vorgenommen wird. Weiterhin deuten unsere Ergebnisse darauf hin, dass sich die relative Fähigkeit von Faktormodellen, Output vorherzusagen, mit zunehmendem Prognosehorizont verschlechtert; hingegen scheint der Prognosehorizont keinen Einfluss auf die relative

Faktorprognosefähigkeit von Inflation zu haben. Andere mögliche Bestimmungsgrößen sind eher irrelevant, etwa ob Prognostiker Faktoren auf Basis eines balancierten oder eines unbalancierten Panel schätzen, ob Restriktionen, die durch die Faktorstruktur impliziert werden, auch in der Prognosegleichung auferlegt werden oder ob jene unrestringiert geschätzt wird und ob iterative oder direkte Mehrschrittprognosen durchgeführt werden. Schließlich finden wir keine Anhaltspunkte dafür, dass eine Vorauswahl der Variablen, die in den Datensatz eingehen, aus dem die Faktoren extrahiert werden, eine Verbesserung der relative Prognosefähigkeit von Faktormodellen bewirkt.

Unsere Analyse kann angewandten Prognostikern helfen, ihre Vorhersage zu verbessern. Unsere wichtigsten Botschaften sind die folgenden. Erstens scheint es sich zu lohnen, Informationen aus großen Datensätzen auszunutzen. Zweitens sollte das Prognosemodell sorgfältig spezifiziert werden. Zwar stellen komplexere Faktorschätztechniken höhere Anforderungen an die Prognostiker im Hinblick auf die Spezifikation, sie schneiden tendenziell aber besser ab, und auch die restringierte Schätzung der Prognosegleichung ist der unrestringierten nicht unterlegen. Drittens deuten unsere Untersuchungen darauf hin, dass einige möglichen Bestimmungsgründe der relativen Prognosefähigkeit von Faktormodellen überraschend irrelevant sind, angesichts der Tatsache, dass Forscher Ihnen in der Vergangenheit relativ viel Aufmerksamkeit beigemessen haben. Dies bezieht sich zum Beispiel darauf, ob und inwiefern vor der Faktorschätzung eine Auswahl der Variablen getroffen, ob durch die Faktorstruktur implizierte Restriktionen auch in der Prognosegleichung berücksichtigt und ob auf ein balanciertes oder ein unbalanciertes Panel abgestellt werden soll. Dagegen hat sich eine andere Bestimmungsgröße als bedeutend und robust herausgestellt, nämlich ob rollierend oder rekursiv prognostiziert wird. Dies ist erstaunlich, da ihr bislang wenig Beachtung geschenkt wurde.

Contents

1	Introduction	1
2	Forecasting with dynamic factor models	3
2.1.	A two-step forecasting approach	4
2.2.	Determinants of factor forecasts	5
3.	Preparing the meta-analysis	12
3.1.	Collecting relevant papers	12
3.2.	Meta-dependent variable	12
3.3.	Meta-independent variables	13
4.	Results	17
4.1.	Descriptive analysis	17
4.2.	Meta-regression	18
4.2.1.	Baseline meta-regression	18
4.2.2.	Robustness checks: outliers, sampling bias, dependency and qualitative differences	21
5.	Conclusion	24
	References	32

Tables and Figures

Table 1	Studies included in the meta-analysis	26
Table 2	Descriptive statistics	28
Table 3	Meta-regression results for output	29
Table 4	Meta-regression results for inflation	30
Figure 1	Smoothed histograms of relative RMSEs	31

How Good are Dynamic Factor Models at Forecasting Output and Inflation? A Meta-Analytic Approach^{*}

1. Introduction

Large-dimensional dynamic factor models are being increasingly applied by policymakers and economic research institutions to forecast key macroeconomic variables, such as real output and inflation. This is partly because many time series are nowadays readily available, and modern computers and software allow us to efficiently summarize the information contained in large datasets. The use of dynamic factor models has been further improved by recent advances in estimation techniques proposed by Stock and Watson (2002a; henceforth SW), Forni, Hallin, Lippi and Reichlin (2005; henceforth FHLR) and Kapetanios and Marcellino (2004; henceforth KM). The two former approaches rely on static and dynamic principal component analysis (PCA), respectively, and the latter on a subspace algorithm. These techniques allow forecasters to easily summarize the information contained in large datasets and extract a few common factors from them. All or a subset of the estimated factors are then entered into rather simple regression models to predict key macroeconomic variables.

Exploiting information from large panels, normally, should help to improve forecasts, and early results were very promising in this respect (cf. SW, FHLR). However, more recent applications find no or only minor improvements (cf. Schumacher 2006; Schumacher and Dreger 2004; Gosselin and Tkacz 2001; Banerjee et al. 2004; Angelini et al. 2001). These conflicting results have launched a lively discussion on whether large-scale factor models are really as useful for forecasting practice as first expected. Some researchers speculate about the conditions under which factor models perform well in forecasting. Banerjee et al. (2005), for example, claim that factor models are relatively good at forecasting real variables in the US but less so in the euro area, whereas, according to the authors of this paper, euro-area nominal variables are easier to predict using factor models than US nominal variables. Likewise, Boivin and Ng (2006) have argued that the composition of the dataset, and not only the pure size of the cross-section dimension, is crucial to producing passable forecasts with factor models. Another debated issue is which of the factor estimation techniques performs best. As will be explained in detail below, the dynamic approaches of FHLR and KM are potentially more efficient than the static SW approach, but also more prone to misspecification.

^{*} Affiliations: Sandra Eickmeier, Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, D – 60431 Frankfurt am Main and University of Cologne, sandra.eickmeier@bundesbank.de; Christina Ziegler, University of Bonn (and visitor of the Deutsche Bundesbank while working on this project), ctiegl@uni-bonn.de.

We are grateful to Jörg Breitung, Ben Craig, Ard den Reijer, Heinz Herrmann, Robinson Kruse, Christian Offermanns and Christian Schumacher for their helpful comments and for the discussions. This paper was presented at a seminar at the Deutsche Bundesbank and at macroeconometric workshops in Berlin and Halle.

The motivation for our paper is twofold. First, a considerable number of papers have been written now, and forecasters from policy institutions are on the verge of integrating factor models into the regular forecasting process. We therefore believe that it is time to systematically summarize this literature. Second, our paper is an attempt to contribute to the discussion on the determinants of the relative forecast performance of factor models. It is unclear *a priori* how some of the potential determinants affect the factor forecast performance, but this question needs to be solved empirically. Our paper tries to identify the relevant determinants and to indicate ways in which factor model forecasts can be improved further.

Our paper surveys studies which have predicted real economic activity and inflation using large-scale dynamic factor models. We summarize the papers' findings and assess which potential determinants of the relative factor forecast performance have been relevant in existing empirical applications. For this purpose, we carry out a meta-analysis. Meta-analyses were first applied in health, educational and psychological sciences for some time, and recently became popular also in macroeconomics (cf. Stanley 2001 and articles in the special edition of the *Journal of Economic Surveys* 2005, Vol. 19(3) and Weichselbaumer and Winter-Ebmer 2005, which, in methodological respects, is closely related to our analysis). They are powerful tools to systematically summarize previous studies' findings, based on formal statistical and econometric techniques, and to detect their determinants. The idea is to collect existing studies which are on a certain issue to be studied, take statistics of interest from these studies, look at empirical distributions and regress these statistics on a number of determining characteristics. Our statistic of interest (or meta-dependent variable) measures the relative factor forecast performance, and we take the root mean squared error (RMSE) of a forecast based on a large-scale dynamic factor model relative to the RMSE of a forecast based on a benchmark model. Overall, we collect 45,111 statistics (or converted related statistics into relative RMSEs, if necessary) out of a total of 46 studies. We provide some descriptive statistics and test in particular, whether factor model forecasts are significantly better or worse compared to other models' forecasts. Theory gives us some guidance on possible determinants of the factor forecast performance. We record these determinants for each observation and regress relative RMSEs on them to examine which forecast environments and designs lend themselves to factor forecasts – and which do not.

Our paper is related to other papers surveying (among others) factor forecast applications such as Reichlin (2003), Stock and Watson (2006) and Breitung and Eickmeier (2006). Those papers all adopt a narrative approach which is more prone to the subjectivity regarding the choice of papers and results. Our paper is also related to a literature which concentrates on certain aspects of factor forecasting in an otherwise broad context. Examples are Kapetanios and Marcellino (2004) and Schumacher (2006) who explicitly compare factor estimation techniques; Boivin and Ng (2005) and D'Agostino and Giannone (2006), who also concentrate on the implementation of estimated factors in the forecasting equation; and

Boivin and Ng (2006), among others, who look at the variables included in the dataset from which the factors are extracted. The advantage of our meta-analytic approach is that many possible determinants, and not just a few, can be considered simultaneously. It has a very broad scope, which allows us to reconcile and explain differences in findings across individual studies.

The results of our analysis can be summarized as follows. On average, factor forecasts are slightly better than other models' forecasts. In particular, factor models tend to outperform small-scale models, whereas they perform slightly worse than alternative methods which are also able to exploit large datasets. Our results further suggest that factor forecasts are better for US than for UK macroeconomic variables, and that they are better for US than for euro-area output; however, there are no significant differences between the relative factor forecast performance for US and euro-area inflation. There is also some evidence that factor models are better suited to predict output at shorter forecast horizons than at longer horizons. These findings all relate to the forecasting environment (which cannot be influenced by the forecasters). Among the variables capturing the forecasting design (which can, by contrast, be influenced by the forecasters), the size of the dataset from which factors are extracted seems to positively affect the relative factor forecast performance. There is some evidence that quarterly data lend themselves better to factor forecasts than monthly data. Rolling forecasts are preferable to recursive forecasts. The factor estimation technique seems to matter as well. Other potential determinants - namely whether forecasters rely on a balanced or an unbalanced panel, whether restrictions implied by the factor structure are imposed in the forecasting equation or not and whether an iterated or a direct multi-step forecast is made - are found to be rather irrelevant. Moreover, we find no evidence that pre-selecting the variables to be included in the panel from which factors are extracted helped to improve factor forecasts in the past.

Our paper is organized as follows. Section 2 presents the approximate dynamic factor model and explains how it is used for forecasting in macroeconomics. It further discusses determinants of factor forecast performance. Section 3 describes the preparatory work in the run-up to the meta-analysis including the collection of relevant papers and the construction of the dataset. Section 4 presents some descriptive statistics, the meta-analytic design and the results. Section 5 concludes.

2. Forecasting with dynamic factor models

We consider a situation where a forecaster is interested in predicting a certain macroeconomic (target) variable y_t . He/she may do this by fitting small-scale time series models such as AR models, which exploit the dynamics of the target variable itself, or VAR models, which, in addition, account for interdependencies among a few variables, to y_t . These simple models have performed fairly well in the past. Nowadays, however, lots of data are available to

forecasters. Those data may contain information which is useful to predict y_t . As has been shown by numerous studies, economic variables strongly comove, and some variables are leading and may provide useful signals for others.

It is, however, not feasible to include all potentially relevant variables simultaneously in a forecasting equation. And this is where factor models come into play. The idea underlying factor models is that the bulk of variation of many variables can be explained by a small number of common factors or shocks. This idea dates back to Burns and Mitchell (1946), and, for instance, Giannone et al. (2004) and Uhlig (2004) have shown empirically that economies are driven by a few factors or shocks. Factor models exploit the variables' comovement and efficiently reduce, in a first step, the dimension of the dataset to just a few underlying factors. In a second step, these factors are included into a rather small forecasting equation to predict y_t , and only a few parameters need to be estimated. Let us explain these two steps in some detail.

2.1. A two-step forecasting approach

This subsection presents the approximate dynamic factor model and how it is used to make forecasts. It is assumed that a large number of variables x_{it} , $i=1,\dots,N$, collected in $X_t = [x_{1t} \dots x_{it} \dots x_{Nt}]$, are driven by few ($q \ll N$) unobserved common factors, summarized in $f_t = [f_{1t} \dots f_{qt}]$. Accordingly, dynamic factor models express the time series x_{it} as the sum of a common component χ_{it} and an idiosyncratic component ξ_{it} . The common component is the product of the $q \times 1$ vectors of dynamic factors which are common to all variables in the set, f_t , (and possibly their lags) and the factor loadings $\lambda_i(L) = \lambda_{i0} + \lambda_{i1}L + \dots + \lambda_{is}L^s$:

$$x_{it} = \chi_{it} + \xi_{it} = \lambda_i(L)'f_t + \xi_{it}. \quad (1)$$

It is useful to write down the static representation of the model:

$$x_{it} = \Lambda_i'F_t + \xi_{it}, \quad (2)$$

where F_t is a vector of $r \geq q$ static factors that comprises the dynamic factors f_t and all lags of the factors which enter with at least one non-zero weight in the factor representation. The $r \times 1$ vector Λ_i comprises all non-zero columns of $(\lambda_{i0}, \dots, \lambda_{is})$. The factors are orthogonal to each other. It is assumed that the factors follow a $\text{VAR}(p^F)$ process and that the idiosyncratic components are weakly serially and cross-correlated in the sense of Bai and Ng (2002). For forecasting purposes, it is also useful to represent the idiosyncratic errors as $\text{AR}(p^\xi)$ processes (cf. Boivin and Ng 2005)

$$(\mathbb{I} - A(L))F_t = u_t, \quad (1 - \rho_i(L))\xi_{it} = v_{it}, \quad (3)$$

where $A(L) = A_1L + A_2L^2 + \dots + A_{p^F}L^{p^F}$ and $\rho_i(L) = \rho_{i1}L + \rho_{i2}L^2 + \dots + \rho_{ip^{\xi}}L^{p^{\xi}}$.

In the literature, there exist basically three methods of estimating the factors F_t from a large dataset, namely those developed by SW, FHLR and KM. We will explain the different estimation methods in detail below, but let us, for the moment, simply assume that \hat{F}_t denotes the vector of estimated factors by one of the three methods.¹

In a second step, the estimated factors or a subset thereof – let us denote them by the $r \times 1$ vector \hat{F}_t – are included in a forecasting equation to predict y_t which may or may not be included in X_t .² The equation is usually given by

$$y_{t+h} = \alpha_0 + \alpha_1(L)' \hat{F}_t + \alpha_2(L)' y_t + \varepsilon_{t+h}, \quad (4)$$

where h is the forecast horizon, $\alpha_i(L) = (\alpha_{i0} + \alpha_{i1}L + \alpha_{i2}L^2 + \dots + \alpha_{ip^{\alpha_i}}L^{p^{\alpha_i}})$, $i = \{1, 2\}$ are r -dimensional ($i = 1$) and scalar ($i = 2$) lag polynomials and $\alpha_i(L)$ can be estimated unrestrictedly or restrictedly. Forecasts are evaluated based on forecast error losses, and forecast error losses obtained from factor model forecasts are then compared with forecast error losses obtained from some benchmark models.

Papers using dynamic factor models to predict macroeconomic variables differ in various respects and, not surprisingly, come to different conclusions regarding the relative forecasting performance of factor models. Theory provides some guidance as to which conditions and what forecast design should lead to good outcomes.

2.2. Determinants of factor forecasts

If the factors and parameters were known and the target variable was correlated with the factors, factor models should deliver smaller forecast errors compared to AR models. This is nicely illustrated in Boivin and Ng (2005) and - to make our paper self-contained - we roughly repeat their illustration in the following paragraph.

Let us suppose for the moment that the variable we want to predict is x_{it+1} and is, thus, contained in X_t , and – for simplicity – that $r = 1$ and the factor and the idiosyncratic component follow AR processes of order 1. Iterating equation (2) one period forward and plugging in expressions for F_{t+1} and ξ_{it+1} , we obtain

$$x_{it+1} = \Lambda_i'(A F_t + u_{t+1}) + \rho_i \xi_{it} + v_{it+1}. \quad (5)$$

¹ Here and in the following, a ‘ $\hat{}$ ’ stands for an estimate.

² It may be that $\hat{F}_t = \hat{F}_t$, i.e. that the factors underlying X_t are identical to the factors included in the forecasting equation. However, some papers include $\hat{F}_t \neq \hat{F}_t$ in the forecasting equation, where \hat{F}_t is a subset of \hat{F}_t (in this case, $r \leq r$). And sometimes r is not estimated and a large number of factors are included in (4).

Solving (2) for ξ_{it} and plugging this expression into (5) yields

$$x_{it+1} = \rho_i x_{it} + \Lambda_i' (\Lambda - \rho_i) F_t + \Lambda_i' u_{it+1} + v_{it+1}. \quad (6)$$

From (6), it is apparent that AR(1) forecasts can be regarded as a special case (for $\Lambda_i = 0$) of factor forecasts. Moreover, factor forecasts deliver smaller mean squared errors than AR(1) models as long as the factors and parameters are known, the target variable is correlated with the factors, i.e. $\Lambda_i \neq 0$, and the factors and idiosyncratic components have different dynamics, i.e. $\Lambda \neq \rho_i$.

While the latter condition is generally fulfilled in practice, this does not hold for the other conditions. Factors are unknown. They need to be estimated and the factor estimates, not the true factors are included in the forecasting equation. Moreover, the target variable is not necessarily correlated with the factors included in the forecast equation, and introducing uninformative factors in the forecast equation would only increase sampling variability. Finally, researchers do not know the models' parameters and are not immune to misspecifications in the forecasting equations.

Whether forecasters face these problems, will depend on the specific forecasting environment and the forecasting design. While they generally have to take the former as given, they can choose the factor forecast design and have, in this respect, some influence on the outcomes. In the following, we identify the determinants of the relative factor forecast performance, classify them into determinants capturing the forecast environment (which cannot be influenced by the forecasters) and those affecting the forecast design (which can be influenced by the forecasters) and discuss implication for the precision of factor estimates, the commonality of the target variable and the specification of the forecasting equation.

Forecasting environment (which cannot be influenced by the forecasters)

The degree of commonality obviously differs across variables, with some variables linked more closely to the overall economic development than others. We will below distinguish forecasts of output and inflation in different countries or regions.

In addition, the relative forecast performance of factor models should vary with the benchmark model, with some models accounting more for the commonality of the target variable than others. While univariate models such as random walks and ARIMA models do not consider the cross-variation between the variables at all, other popular benchmarks such as VAR models or single equation models with one or a very few observable indicators do; also, the issue of whether the target variable is related to a stronger extent to the factors or to the observable indicators will govern the relative forecast performance. Recently, alternative models which are able to exploit the data-rich environment, such as forecast combination, model averaging including Bayesian model averaging and bagging, ridge regression,

shrinkage and partial least squares have been proposed (cf. Stock and Watson 2004 and 2006, Lin and Tsay 2005).

Also, the forecast horizon is generally given to the forecasters. Two arguments come to our mind. The greater the predictability of the common component relative to the idiosyncratic component at larger horizons (which is positively related to the relative persistence of the two components), the better the forecast performance of factor models should be relative to small-scale models such as AR models. Moreover, if X_t contains leading indicators of y_t , which is typically the case, we would expect factor models to be more successful than, for instance, univariate models at predicting y_t at longer horizons.

The commonality of the target variable and, hence, the relative forecast performance of factor models could also be altered by the specific estimation period. Variables comove more in certain periods than in others, and we would therefore expect factor models to be better suited for forecasting in periods of greater comovements. It is, for example, well known that variables move relatively closely together during economic downturns. Likewise, periods which are characterized by lower economic volatility, such as the great moderation period since 1985, are less well described by a factor model, as shown by D'Agostino and Giannone (2006). We are not able to capture this effect here, since only very few papers consider periods which are clearly characterized as being periods of either high or low comovement.

Forecasting design (which can be influenced by the forecasters)

Choosing the forecasting design mainly means constructing the dataset from which the factor are extracted and estimating the factors and the forecasting equation.

Regarding the dataset, it is a well-known result of the factor literature that there are benefits from information contained in large datasets in terms of greater precision of the factor estimates. Stock and Watson (2002a) and Bai and Ng (2002), for example, show that the uncertainty associated with the factor estimation becomes negligible and factors can be treated as known if the cross-section dimension of X_t , N , and the number of observations, T , tend to infinity. This result suggests that the forecast performance of factor models should improve with N and T .

However, various studies demonstrate that it is not the pure size of the dataset,³ but also its characteristics, which matter for forecasting. Forecasters should take care that variables which are highly correlated among each other are included in the dataset; this should improve the precision of the factor estimation. Moreover, the dataset should contain variables which are

³ Watson (2003) and Bai and Ng (2002) show in real-time experiments and in simulations that there are basically no gains from increasing N beyond 50 or 40, respectively. Boivin and Ng (2006) demonstrate that increasing N beyond a certain number can even be harmful and may result in efficiency losses.

highly correlated with the target variable. Forecasters often face a trade-off between including many time series and/or many observations and time series which satisfy these two requirements. This trade-off is reflected in the following determinants.

One issue is whether it is well suited to extract factors from a balanced or an unbalanced panel. Forecasters relying on an unbalanced panel argue that additional information from time series with missing observations can be exploited. Improvements in terms of forecasting performance can, however, only be expected if these additional time series comove with the other variables contained in the panel and/or with the target variable.

Another issue is raised by Boivin and Ng (2006). They show that the inclusion of variables with errors which have large variances and/or are cross-correlated, should worsen the precision of factors estimates. The forecast performance of factor models may also be worsened if variables that are irrelevant for y_t are included in X_t - this is referred to as the oversampling problem. Boivin and Ng (2006) suggest pre-selecting the variables to be included in X_t and removing variables with correlated and/or large errors and/or variables which are irrelevant for the target variable prior to estimating the factors.

A third issue which is relevant in this context is whether to use a recursive or a rolling factor estimation (and forecasting) scheme. A forecast based on a recursive scheme relies on an estimation period of increasing length, where the starting point remains fixed, whereas a rolling scheme relies on a fixed-length window which is shifted every period. A priori, it is unclear which scheme would yield more precise factor estimates. On the one hand, the recursive scheme allows us to exploit more information since estimation tends to be based on a larger T . On the other hand, if factors (and/or factor loadings) are subject to structural breaks which occur in the course of the estimation period, a rolling scheme which gives observations in the past a lower weight compared to a recursive scheme should deliver better forecasts.

A fourth issue regards the frequency. Quarterly time series correspond to averages of the monthly series or smoothed monthly series. If idiosyncratic noise is averaged away, as one might expect, rather than the common part of the variables, commonality will be greater at quarterly than at monthly frequency, although monthly data potentially contain more information.

Besides the size and the composition of the dataset, estimation techniques also matter for factor forecasts. This refers to the factor estimation technique and to the specification and estimation of the forecasting equation.

The technique used to estimate the factors should affect the precision of the estimates. As already pointed out above, basically three different methods are employed in the literature, namely those proposed by SW and FHLR and the more recently developed, still less frequently applied KM method. Let us briefly explain them.

SW propose estimating F_t with static PCA applied to X_t . The factor estimates are simply the first r principal components of X_t , $\hat{F}_t^{SW} = \hat{\Lambda}' X_t$, where $\hat{\Lambda}$ is the $N \times r$ matrix of the eigenvectors corresponding to the r largest eigenvalues of the sample covariance matrix $\hat{\Sigma}$.

FHLR propose a weighted version of the principal components estimator suggested by SW, where time series are weighted according to their signal-to-noise ratio, which is estimated in the frequency domain. The authors proceed in two steps. First, the covariance matrices of common and idiosyncratic components of X_t are estimated with dynamic PCA. This involves estimating the spectral density matrix of X_t , $\Sigma(\omega)$, which has rank q . For each frequency ω , the largest q eigenvalues and the corresponding eigenvectors of $\Sigma(\omega)$ are computed, and the spectral density matrix of the common components $\hat{\Sigma}_\chi(\omega)$ is estimated. The spectral density matrix of the idiosyncratic components is given by $\hat{\Sigma}_\xi(\omega) = \hat{\Sigma}(\omega) - \hat{\Sigma}_\chi(\omega)$. Inverse Fourier transform provides the time-domain autocovariances of the common and the idiosyncratic components $\hat{\Gamma}_\chi(k)$ and $\hat{\Gamma}_\xi(k)$ for lag k . Since dynamic PCA corresponds to a two-sided filter of the time series, this approach alone is not suited for forecasting. Therefore, FHLR, in a second step, search for the r linear combinations of X_t that maximize the contemporaneous covariance explained by the common factors $\hat{Z}_j' \hat{\Gamma}_\chi(0) \hat{Z}_j$, $j = 1, \dots, r$. This optimization problem is subject to the normalization $\hat{Z}_j' \hat{\Gamma}_\xi(0) \hat{Z}_i = 1$ for $i = j$ and 0 for $i \neq j$. It can be reformulated as the generalized eigenvalue problem $\hat{\Gamma}_\chi(0) \hat{Z}_j = \hat{\mu}_j \hat{\Gamma}_\xi(0) \hat{Z}_j$, where $\hat{\mu}_j$ denotes the j -th generalized eigenvalue and \hat{Z}_j its $N \times 1$ corresponding eigenvector. The factor estimates are obtained as $\hat{F}_t^{FHLR} = \hat{Z}' X_t$ with $\hat{Z} = [\hat{Z}_1 \dots \hat{Z}_r]$.

KM propose a state-space framework to estimate the factors. The starting point is the prediction error representation $X_t = \Lambda F_t + \Xi_t$, $F_{t+1} = \Lambda F_t + L \Xi_t$, where Ξ_t is the vector of innovations. KM apply a subspace algorithm which allows the factors to be estimated without specifying and identifying the full state space model. The model can be written as a vector equation $X_t^f = OKX_t^p + E\Xi_t^f$. $X_t^f = [X_t' \ X_{t+1}' \ \dots]$, $X_t^p = [X_{t-1}' \ X_{t-2}' \ \dots]$, where, in practice, leads and lags need to be truncated, $\Xi_t^f = [\Xi_t' \ \Xi_{t-1}' \ \dots]$ and the matrices O , K and E are complicated functions of the parameters in the prediction error representation. OK is estimated as $(X_t^p' X_t^p)^+ X_t^p' X_t^f$, where B^+ denotes the Moore-Penrose inverse of a matrix B . The coefficient can be decomposed by a singular value decomposition $(X_t^p' X_t^p)^+ X_t^p' X_t^f = \hat{U} \hat{S} \hat{V}'$. The factor estimates are given by $F_t^{KM} = \hat{K} X_t^p$, where $\hat{K}_t = \hat{U}_r \hat{S}_r^{1/2}$, and \hat{U}_r denotes the first r columns of the left singular value matrix \hat{U} , and $\hat{S}_r^{1/2}$ is the $r \times r$ upper left square matrix of the square root of the singular value matrix \hat{S} containing the largest singular values in descending order.

It is not clear *a priori* which of the three methods will perform best in practice. Weighting the time series according to their signal-to-noise ratios, as is done in FHLR, should deliver efficiency gains compared to the unweighted SW and KM versions. Efficiency gains should also be obtained because the FHLR and KM methods allow for richer dynamics: factors are estimated as linear combinations of contemporaneous time series and their leads and lags,

whereas only contemporaneous comovement between variables is accounted for in the approach originally proposed by SW.⁴ The SW approach, by contrast, has the advantage that it only requires the estimation of a single auxiliary parameter (r), whereas more unknown parameters have to be set in KM and FHLR,⁵ thus making the latter approaches more vulnerable to misspecification. Also, if no lagged relationship between x_{it} and F_t exists in the data, unnecessary estimation of the spectral density matrix for the FHLR approach could induce efficiency losses (cf. Bai and Ng 2005). The KM approach clearly gives more structure to the data than SW and accomplishes this - at least in existing practical applications - by the rather restrictive processes assumed for the innovations and the factors.⁶ This may be more efficient if the data are well described by this structure. However, overly tight restrictions will lead to less precise KM factor estimates.

It has recently been pointed out by Boivin and Ng (2005) and discussed by D'Agostino and Giannone (2006) that the approaches of SW and FHLR - they do not consider the KM approach - differ not only in the way the factors are estimated, but also in another respect. FHLR impose the restrictions implied by the factor model (1) in the forecasting equation (4), i.e. $\alpha_1(L)$ is a function of the loadings associated with y_t which, in this case, needs to be included in X_t and the dynamics of the factors and the idiosyncratic components and $\alpha_2(L)$ depends on the latter dynamics.⁷ By contrast, SW propose estimating the forecast equation (4) unrestrictedly with OLS. Again, the impact of imposing or not imposing the restrictions implied by the factor model is unclear and depends strongly on whether or not the factor model is correctly specified and the parameters associated to it are precisely estimated.

The forecast performance of factor models could also be affected by whether direct or iterated multi-step forecasts are made. Iterated forecasts use a one-period ahead model, iterated forward for h periods, whereas direct forecasts use a horizon-specific estimated model where the dependent variable is the multi-period ahead value being forecasted (Marcellino et al. 2006). Theoretically, iterated forecasts are more efficient if the one-period model is correctly specified, but direct forecasts are more robust to model misspecification. In our context of dynamic factor forecasts, this means that iterated forecasts will yield improvements over direct forecasts in terms of smaller relative forecast error losses of factor models if the VAR model of the factors, i.e. the first part of (3), is correctly specified.⁸

⁴ An exception is the “stacked” version of the SW method, where factors are estimated as linear combinations of X_t and its lags. This approach is also used by Grenouilleau (2004).

⁵ Besides r , the number of dynamic factors q , the truncation lag parameters for spectral estimation as well as the number of frequency grids needs to be chosen in FHLR. The KM model requires to set r as well as the truncation leads and lags for the subspace algorithm.

⁶ Innovations are assumed to be serially uncorrelated. Factors are generally assumed to follow a VAR(1) process.

⁷ The nonparametric forecast of FHLR involves predicting the common component as $\hat{\chi}_{t+h} = X_t \hat{Z} (\hat{Z}' \hat{\Sigma} \hat{Z})^{-1} \hat{Z}' \hat{\Gamma}_x(\omega)$ and, thus, takes the restrictions implied by the factor model into account.

⁸ Some forecasters, in addition, consider forecasts of the idiosyncratic components (cf. den Reijer 2005). In this case, it will also matter if the AR model of the idiosyncratic component, i.e. the second part of (3), is correctly specified.

Factor forecast applications also differ in other respects concerning the specification of the forecasting equation (4), in particular in their choice of factors and lags of the factors and dependent variables. Some determine r based on formal information criteria or on other, rather informal, criteria⁹ and include all estimated factors ($\underline{r} = r$) in equation (4) (cf. Schneider and Spitzer 2004, Schumacher 2006, Van Nieuwenhuyze 2006). Others include only the first factor in the forecasting equation which is often seen as a measure of the business cycle (see, for example Watson 2003 or the CFNAI constructed by the Chicago Fed¹⁰) or core inflation (cf. Camba-Méndez and Kapetanios 2005). Some also consider the first $\underline{r} > 1$ factors, where \underline{r} is chosen somewhat *ad hoc* (Lin and Tsay 2005, Stavrev 2006). Bruneau et al. (2003b) include the first, the second, etc., each at a time, in equation (4) to assess the marginal contribution of each of the factors to the forecast. Most papers, however, set a maximum number of factors and lags (of the dependent variable and/or the factors) and determine the factors and the lags to be included in the forecasting equation simultaneously using Akaike or Bayesian information criteria (cf. Matheson 2006, Stock and Watson 1999, 2003, Banerjee et al. 2005 and 2006, Artis et al. 2004, Jeon 2004) or performance-based measures such as the mean squared error (cf. Schumacher 2006, Forni et al. 2001 and 2003). Others do not consider lags of the factors (Schumacher 2006, Giacomini and White 2006, Forni et al. 2001, Boivin and Ng 2006) and/or autoregressive terms from the outset (cf. Liu 2004, Stock and Watson 1998, Tatiwa Ferreira et al. 2005). This presentation clearly shows that there is much heterogeneity among the papers with respect to the factors and lags included in the forecasting equation. Since it would be difficult to classify results meaningfully, we do not address this issue here.

These considerations help us to choose our meta-independent variables. To summarize: We will consider variables reflecting the forecasting environment (which are given to the forecaster), such as the variable itself, the benchmark model and the forecast horizon. In addition, our set of meta-independent variables comprises variables capturing the forecasting design (which can be influenced by the forecaster), such as N and T ; whether forecasters rely on balanced or unbalanced panels; whether the variables to be included in X_t have been pre-selected prior to estimating the factors; the frequency; whether a recursive or a rolling forecasting scheme was employed; which techniques were used as the basis for estimating the factors; variables which state whether the restrictions implied by the factor model are imposed in the forecasting equation and whether iterated or direct multi-step forecasts were performed. Details on these and other meta-independent variables are presented in the next section.

⁹ Formal criteria to determine r are those by Bai and Ng (2002). Examples of informal criteria are those developed by FHLR to determine q or the variance share explained by the static factors to determine r .

¹⁰ http://www.chicagofed.org/economic_research_and_data/cfnai.cfm.

3. Preparing the meta-analysis

Before the meta-analysis can be carried out, much preparatory work needs to be done. Relevant papers have to be collected, the independent and dependent meta-variables covering the relative forecast performance of dynamic factor models and its determinants, respectively, have to be chosen, and, based on this decision, the dataset has to be constructed. Replicability and completeness are important principles in meta-analysis, which we therefore try to follow (Lipsey and Wilson 2000, Stanley 2001).

3.1. Collecting relevant papers

We start with an extensive computer search in the EconLit, Google scholar and IDEAS databases. We look for empirical studies on macroeconomic forecasting with factor models. The keyword is “forecast” combined with “factor models”, “dynamic factors” “principal components” or “diffusion index”. We further search in the working papers series of central banks, the Bank for International Settlements, and the International Monetary Fund, and look at the websites of researchers who are known in the research community for being specialized in the field of dynamic factor modelling and forecasting. Some papers’ main focus is the forecast performance of factor models, while others only use them as benchmarks. To focus our paper, we concentrate on studies that forecast real economic activity and inflation.¹¹

We include in our sample published as well as unpublished papers which comprise working papers and manuscripts, so that we can consider as many results as possible. Forecasting with factor models is a relatively new field of research. This is reflected in the fact that only 52 percent of the papers we consider are already published (or forthcoming) and that most unpublished papers were written up to two or three years ago. Unpublished paper versions of the published papers are generally also available to use. In one case, the unpublished version provides more results than the published version, which has been shortened for publication. In this case, we consider all results reported in the published version plus those in the unpublished version which have not already been taken from the published version.¹² Overall, we rely on 46 studies which are listed in Table 1, with published and unpublished versions being counted as one paper.

3.2. Meta-dependent variable

An important decision to be made is on the dependent variable of our meta-regression. This variable is supposed to measure the forecast performance of a factor model relative to some

¹¹ There are papers which also predict monetary policy and financial variables using factor models. The corresponding results are, however, not considered here.

¹² Fidrmuc and Korhonen (2006) also consider both, working paper and published versions in their meta-analysis.

benchmark model. Such a measure is ideally contained in all collected papers. Most studies report forecast error losses such as mean squared errors (MSE) or root mean squared errors (RMSE) of models used in these studies to predict a certain target variable or forecast error losses of more complex models relative to simple benchmark models such as random walks or AR models. We decide to focus in our analysis on the RMSE of factor models relative to the RMSE of a certain benchmark model. Let us denote this ratio associated to observation $j = 1, \dots, \tilde{n}$ as ψ_j ;

$$\psi_j = \left(\frac{RMSE^{DFM}}{RMSE^{Bench}} \right)_j, \quad (7)$$

where an observation refers to a result in a study and benchmark models can differ across papers. Let us summarize all observations in $\tilde{\Psi} = [\psi_1 \dots \psi_j \dots \psi_{\tilde{n}}]$, where \tilde{n} equals the total number of observations, which is 45,111. Results in the papers that were not already defined as in (7) were converted. Notice that we tried to include all results in order to avoid any selection bias.

$\tilde{\Psi}$ contains some large outliers. As shown by Rousseuw and Leroy (1987), outliers can distort parameter estimates. We therefore remove observations outside 1.5 times the interquartile range. As we will show, the results are relatively robust with respect to other outlier correction methods. After outlier removal, we are left with $n = 38,729$ observations, 18,910 of which refer to output and 19,819 to inflation. Let us denote the outlier-adjusted $n \times 1$ vector of relative RMSEs by Ψ .

3.3. Meta-independent variables

As discussed in the previous section, theory gives us some guidance as to which variables may determine the forecast performance of factor models. In addition to these variables, we consider variables capturing the publication strategy. In the following, we list and briefly explain the meta-independent variables. Some of them are continuous variables, others discrete. The latter can be divided into certain cases which are given in parentheses.

- VARIABLE (Y, INFL) captures the economic meaning of the target variable y_t . We distinguish between variables covering real economic activity (Y) and those covering inflation (INFL). Real economic activity includes GDP, industrial production, employed persons, hours and unemployment, retail sales, real personal income, real manufacturing trade and sales, consumption, investment, inventories and orders (in levels or first differences). Inflation measures are consumer prices, producer prices, retail prices and other sub-aggregates, wages, measures of core inflation and the GDP deflator (in first or second differences).
- COUNTRY (US, UK, EA, OTHER) is divided into four groups of predictions: those associated with the US, the UK, the euro area as an aggregate plus individual euro-area

countries (EA) and other countries (OTHER). The latter group contains results for Canada, New Zealand, Japan, Brazil, Chile and central and east European countries.

- BENCH (ARIMA, RW, VAR, INDIC, LARGE) captures the benchmark models to which large factor models are compared. We distinguish between random walks (RW), ARIMA models where most often AR models are employed, VAR models and single equation models with indicators (INDIC) which is similar to equation (4), where \hat{E}_t is replaced with one or more measurable indicators. Notice that INDIC comprises also some structural models such as the Phillips curve, which is often used to predict inflation.¹³ More recently, large-scale factor models are also compared with other models suited to exploit data-rich environments. This class of predictors includes model averaging, forecast combination, ridge regression and partial least squares. We summarize results from these approaches in LARGE.
- HORIZON refers to the forecast horizon. Months were converted to quarters.
- N, the dimension of the cross-section (in logs).
- T, the time dimension of the sample on which the estimation is based (in logs). Notice that, when a recursive forecasting scheme is applied, T varies over time. In this case, we compute the average T, given by $\min\{T\} + (\max\{T\} - \min\{T\}) / 2$.
- BAL (YES, NO) reflects whether the factors are estimated from a balanced (YES) or an unbalanced panel (NO).
- PRESEL (NO, 1, 2). This variable distinguishes whether authors use all the data they have collected to extract the factors (NO) or make a pre-selection, either by removing data with correlated errors and/or errors that have a large variance (1) or by removing variables which they think are irrelevant for the target variable or including potentially relevant subgroups of variables which are formed based on economic considerations (2). Case 1 contains observations recorded from papers such as Boivin and Ng (2006), Schneider and Spitzer (2004), Den Reijer (2005), Van Nieuwenhuyze (2006), Matheson (2006), and Bruneau et al. (2003b), with the first four papers focusing on output and the first and the last two papers on inflation. Boivin and Ng (2006) excludes series with large and correlated errors from the dataset. Schneider and Spitzer (2004), Den Reijer (2005) and Matheson (2006) order the variables according to their correlation with the target variables; Schneider and Spitzer (2004) and Den Reijer (2005) then sequentially include them into the dataset to minimize the forecast error loss, and Matheson (2006) adopts a *ad hoc* approach by including the first 5, 10 and 50 percent in X_t . An *ad hoc* approach is also taken by Van Nieuwenhuyze (2006) who considers only the 75 percent of the variables with the highest commonality ratio. As an alternative, he also selects a dataset

¹³ Other structural models include inflation indicators derived from an SVAR and a Blanchard-Quah decomposition and from a P* model into the forecasting equation (Stavrev 2006).

which maximizes the commonality ratio of the target variable. Bruneau et al. (2003b), finally, estimate their factor model based on a dataset which includes those indicators which, individually, delivered MSEs relative to AR-MSEs significantly below 1. Case 2 includes observations from papers such as Angelini et al. (2001) who extract factors from a nominal and a non-nominal dataset separately to forecast inflation, Bruneau et al. (2003a) who use pure French and Belgian factors to predict euro-area inflation, FHLR who investigate whether financial factors help to predict output and inflation, Inoue and Kilian (2005) who assess whether real variables help predicting inflation, etc. Those papers not only aim at solving the oversampling problem and removing variables which are irrelevant for the target variables, but some of them also focus on the predictive content of certain groups of variables irrespective of the overall forecast performance of the factor model. Unfortunately, these two aspects cannot be regarded separately. This needs to be kept in mind when interpreting results. Notice also that, in a way, some (at least implicit) form of pre-selection always takes place when forecasters construct their large datasets. However, only when variables are included or excluded from the original dataset after applying some formal or explicitly specified criteria, we attribute the resulting observations to cases 1 or 2, otherwise to the case NO.

- ROLREC (ROL, REC) captures whether a rolling (ROL) or a recursive (REC) forecasting scheme is adopted.
- FREQ (Q, M) captures whether the forecast is made on a monthly (M) or a quarterly (Q) basis.¹⁴
- FACTOR (SW, FHLR, KM) distinguishes between the different factor estimation techniques.
- RESTR (YES, NO) captures whether restrictions implied by the factor model are imposed in the forecasting equation (YES) or not (NO).
- ITDIR (IT, DIR, 1_STEP) states whether a direct (DIR), an iterated (IT) multi-step forecast or a one-step ahead forecast (1_STEP) is made.
- PUBL (YES, NO) reflects whether a paper is already published or still a working paper or a manuscript. In meta-analyses, this variable generally captures possible publication bias, where journals' editors have a tendency to publish significant results. It is not totally clear how this translates to our context, but it could be supposed that editors might have a tendency to publish results favoring factor forecasts, at least if factor forecasts are the focus of the paper which is the case for most studies (cf. Stanley 2005, Doucouliagos 2005). Some meta-studies also use this variable to capture differences in quality and weight observations accordingly (cf. Weichselbaumer and Winter-Ebmer 2005). These

¹⁴ SW also show how datasets with mixed frequencies can be exploited. Schumacher and Breitung (2006) use this method to predict German output. Their results are, however, not considered here.

studies presume that published studies which have gone through a rigorous referee process are qualitatively better in the sense that errors are eliminated and only accurate and robust results survive. We are, however, skeptical that this argument applies to our context. As pointed out above, factor forecasting is a relatively new field of research and many papers have simply not yet been published due to long publication lags. Although the interpretation of PUBL is not fully clear, we will keep this variable in our regression.

- AUTHOR (YES, NO) captures if one (or more) of the authors of the particular study was (were) among the developers of (one of) the dynamic factor model(s) used in that particular study, namely Stock, Watson, Forni, Hallin, Lippi, Reichlin, Kapetanios and Marcellino. The hypothesis we test is whether results are biased in favor of factor models when produced by the developers of the model who may be interested in seeing their models widely applied. Whenever authors not only focus on one factor estimation technique but compare different factor estimation techniques and developers of one the applied models are among the authors, only observations that refer to the model which was also developed by (one of) the author(s) are attributed to YES. Observations associated with other factor estimation techniques are associated to NO.

Three remarks are in order. First, there is certainly some overlap between the meta-independent variables. For example, FACTOR_FHLR implies a weighting of time series where weights are inversely related to the variance of idiosyncratic components. This idea is also captured by PRESEL_1, where data with important idiosyncratic components are either downweighted or dropped (cf. Stock and Watson 2006, D'Agostino and Giannone 2006). In order to disentangle the two variables, we include in PRESEL_1 only cases where data are completely eliminated from the dataset, leading thus to weights of 0 or 1. Another difficulty arises, for example, when disentangling different benchmark models. In most cases, there is some pre-testing when deciding which indicators to include in BENCH_INDIC. If indicators are chosen out of a very large set of variables, this also means that information from a data-rich environment is exploited. In fact, Stock and Watson (2005) and Lin and Tsay (2005) attribute single equations models with indicators where the latter are selected from a large dataset to the class of large predictors, and we could have included parts of these models in BENCH_LARGE as well. Our choice is certainly somewhat *ad hoc*. A final example is an overlap between HORIZON and ITDIR where we distinguish between one-step ahead (ITDIR_1_STEP) and multi-step forecasts (ITDIR_IT and IT_DIR); this, however, is inevitable if we want to assess the impacts of iterated and direct forecasts and given that those only refer to multi-step predictions.

Second, more recently, attempts have been made to pool factor model forecasts. Artis et al. (2004) compute simple averages of factor-based forecasts. Inoue and Kilian (2005) consider bagging of factor forecasts, Stock and Watson (2006) and Koop and Potter (2004) focus on Bayesian model averaging, where the models considered are based on principal

components. Pooling factor model forecasts goes beyond our analysis, but it seems promising and could be treated in future work.

Third, although important for readers who aim at understanding and perhaps even replicating results, the designs of the analyses are sometimes insufficiently documented. Whenever some of the characteristics we consider in our meta-analysis were missing from a paper, we sent an e-mail to (one of) the author(s). Although the authors were generally very helpful and we obtained responses very quickly, we were not able to fill all the gaps. Whenever an observation could not be related to the characteristics used as meta-independent variables, we used this observation for the descriptive analysis but had to exclude it from the meta-regression analysis. Overall, the baseline meta-regressions for output and inflation are based on 17,222 and 17,294 observations, respectively.

4. Results

4.1. Descriptive analysis

This subsection presents descriptive statistics of relative RMSEs obtained from factor forecasts associated with the total sample (after outliers were removed) and, separately for output and inflation, the different countries/country groups and benchmark models. We are particularly interested in whether factor models are, on average, better or worse than certain benchmark models. We test whether means and medians of relative RMSEs differ significantly from 1 using a *t*-test and a Wilcoxon sign rank test, respectively. Empirical distributions for the entire sample and for output and inflation separately are shown in Figure 1. Descriptive statistics are provided in Table 2.

The means and the medians for the entire sample and for inflation are roughly at 0.98 and those for output just under 0.99. Although these numbers are only slightly below 1, the tests indicate that, on average, factors models perform significantly better than the respective benchmark models in predicting output and inflation. However, this is associated with more mass in the tails of the empirical distribution of inflation compared to output, as is apparent from the lower panel of Figure 1.

When looking separately at relative RMSEs of different benchmark models, it turns out that factor models perform worse relative to alternative models which are able to exploit large datasets. By contrast, factor models generally outperform small-scale models, with a few exceptions: factor models do worse than ARIMA models and do not significantly differ from VAR models, when predicting inflation. Factor models show the greatest improvements over random walks. It is also notable that single-equation models with indicators perform relatively badly, for instance compared to more simple ARIMA models.

Descriptive statistics for individual countries/country groups clearly indicate that factor models do better than other models on average for the US but worse for the group of other countries (OTHER). This holds for both output and inflation. Results for the UK are not the same for output and inflation; the relative RMSE does not differ significantly from 1 for output but exceeds 1 for inflation. Also interestingly, factor forecasts of euro-area output are outperformed, although not significantly, on average by other models' forecasts, but factor forecasts of euro-area inflation are clearly superior to other models' forecasts and the relative RMSE associated to euro-area inflation is even lower than the relative RMSE associated to US inflation. Our findings therefore support the claim of Banerjee et al. (2005) that factor models are better at predicting nominal variables in the euro area compared to the US and real variables in the US compared to the euro area.

The descriptive analysis masks the fact that the various meta-independent variables may interfere with one another. To disentangle the effects, a regression approach is adopted in the next subsection.

4.2. Meta-regression

4.2.1. Baseline meta-regression

We establish the meta-regression equation

$$\psi_j = \mu + \phi' M_j + \eta_j, \quad (8)$$

where M_j is the vector of explanatory (meta-independent) variables associated with observation j , ϕ is the corresponding vector of coefficients and μ refers to the overall constant. M_j comprises the continuous variables N, T and HORIZON (as deviations from their arithmetic means) as well as a set of dummy variables into which the discrete variables were transformed. Consider, for example, the variable FACTOR. The dummy variables for the cases FHLR and KM take values of 1 if F_j was estimated with the FHLR and the KM technique, respectively, and 0 otherwise. To avoid perfect collinearity, the SW case is left out. Negative/positive signs of the coefficients of the included dummies indicate lower/higher relative RMSEs, i.e. a better/worse relative factor forecast performance, compared to the cases which were left out. The impacts of the cases which are left out are summarized in the common intercept, which, hence, can be interpreted as the average relative RMSE conditional on the characteristics given by the cases which were left out and on the means of the continuous variables.¹⁵ We estimate equation (8) separately for output and inflation, which is

¹⁵ The constant and the variables' coefficients can then be used to compute the means of relative RMSEs conditional on any characteristics, the reader might be interested in.

suggested by an F-test.¹⁶ In our baseline regression, we use OLS and report White-corrected standard errors to account for heteroskedasticity in the errors.¹⁷

The results of our baseline regression equations for output and inflation are given in the first four columns of Tables 3 and 4. While specification (1) refers to a one-time estimation of equation (8), specification (2) provides estimation results where insignificant variables are sequentially removed from the set of meta-independent variables¹⁸. Variables remaining in the equation in (2) have coefficients that are very similar to the corresponding coefficients in (1).

As regards the variables capturing the forecasting environment, it turns out that, consistent with the descriptive statistics, factor models are significantly better at predicting US than British variables. Factor forecasts of euro-area output are clearly outperformed by factor forecasts of US output, yet differences between the US and the euro area do not differ significantly for inflation.

The coefficients of the variables capturing the different benchmark models suggest, in line with the descriptive statistics of the previous section, two conclusions. First, it seems worthwhile to exploit information contained in a data-rich environment. This is derived from our finding that factor models perform relatively better in comparison to small-scale models which do not or barely account for the co-variation between variables than in comparison to alternative predictors suited to handle large datasets. Second, among the small-scale models, ARIMA models perform the best.

While the horizon does not enter the equation for inflation significantly, factor models seem to perform better at shorter than at longer horizons for output.¹⁹ This is surprising given results by Giannone et al. (2002) for the US and Altissimo et al. (2001) for the euro area, who find that the common component of output is much more persistent than its idiosyncratic component.

As expected, the size of the dataset seems to matter, with N and T having favorable significant effects on the relative forecast performance of factor models. Whether forecasters

¹⁶ The F-test is constructed by augmenting (8) with interaction terms: $\psi_j = \mu + \phi' M_j + \phi'(D_j \tilde{M}_j) + \eta_j$, where D_j is the scalar dummy variable for output and \tilde{M}_j corresponds to M_j , but does not contain the dummy for inflation (which is included in M_j). We estimate this equation and test the null that all elements of ϕ equal 0. i.e. that (8) can be applied to the entire sample. We obtain $F(J, n - \tau) = 29.815$ with degrees of freedom $J = 22$ and $n - \tau = 34,470$, where τ is the number of freely estimated parameters, and therefore strongly reject the null.

¹⁷ We also estimated (8) with $\log(\psi_i)$ instead of ψ_i on the left hand side to allow for negative values of the dependent variable. Results remained unaffected. We are also concerned with multi-collinearity and removed variables contained in M_j one by one. The coefficients and the significance level of the remaining variables remain basically the same.

¹⁸ The variable with the lowest t -statistic is removed from the set of meta-independent variables, and relative RMSEs are, again, regressed on the reduced set of meta-independent variables. This is repeated as long as only variables which are significant at the 5 percent level are left in the equation.

¹⁹ This finding still holds when we remove individual variables one by one from the set of variables (including ITDIR_1_STEP which may be somewhat correlated with HORIZON, as pointed out in the previous section). The positive and significant coefficients also survives when we regress relative RMSEs on a constant and HORIZON alone.

rely on a balanced or an unbalanced panel does not seem to matter for the relative forecast performance of factor models. Results regarding the variable PRESEL are ambiguous. While neither the removal of variables with correlated errors and/or errors with large variances (PRESEL_1) nor the removal of variables based on economic considerations (PRESEL_2) affects the forecast performance of factor models significantly in the case of output, both PRESEL_1 and PRESEL_2 enter the inflation equation positively and significantly. This latter finding is difficult to interpret. The positive sign of PRESEL_2 could be explained by the fact that not only papers which try to solve the oversampling problem discussed above, but also papers which aim at assessing the predictive content of certain groups of variables, yet whose main focus is not to improve the overall forecast performance of the factor model fall in this category. However, the positive sign of PRESEL_1 is puzzling.²⁰

A result which holds for both output and inflation is that forecasts based on a rolling forecasting scheme perform better than forecasts based on a recursive scheme. In addition, the effect is quite strong economically: using a rolling forecasting scheme instead of a recursive scheme lowers the relative RMSE of output and inflation by 10 and 16 percentage points, respectively. This suggests that structural breaks in the factors (and/or factor loadings) may be present in most of the datasets used by the forecasters, and accounting for them by using a rolling scheme seems to overcompensate the gains from using long time series.

Another result is that factor models perform relatively well when forecasters rely on quarterly rather than monthly data, although *FREQ_Q* is insignificant for output. For inflation, relative RMSEs are by 10 (or 11 in specification (2)) percentage points lower when forecasts are made on a quarterly rather than on a monthly basis. This confirms our conjecture that idiosyncratic noise rather than common fluctuations are smoothed away when converting monthly series to quarterly series.

Another strong result is that the FHLR and KM methods outperform the SW method. On average, the SW method yields relative RMSEs which exceed relative RMSEs produced based on the FHLR and KM techniques by 8 and 18 percentage points for output and 12 and 3 (or 5 in specification (2)) percentage points for inflation. We carefully conclude that it seems worthwhile accounting for the dynamic relationships between the variables. These results, in addition, suggest that efficiency gains which may result in more precise factor estimation can be obtained from weighting time series according to their signal-to-noise ratios. Our results are in line with FHLR (2003) who find that their estimation approach outperforms the SW approach and also Schumacher (2006) who finds modest improvements of the FHLR and the KM estimation techniques over the SW technique, while, according to Boivin and Ng (2005), D'Agostino and Giannone (2006) and Stock and Watson (2005), there are only minor and

²⁰ The positive sign of PRESEL_1 for inflation coincides with the impression we got from the majority of the (only three) papers, which pre-select the dataset before predicting inflation. Interestingly, when among those papers only the Boivin and Ng (2006) paper is left in the dataset, the coefficient of PRESEL_1 turns (significantly) negative.

unsystematic differences between SW and FHLR.²¹ Our results are especially notable given that, in practice, most forecasters apply the SW method which is much easier to implement.

Our findings also suggest that imposing the restrictions implied by the factor structure in the forecasting equation does not harm the relative factor forecast performance. This supports the results obtained by D’Agostino and Giannone (2006) rather than those of Boivin and Ng (2005) who opt for the more robust unrestricted estimation. We also find that it barely matters from an economic and a statistical point of view whether iterated or direct multi-step forecasts are made. This is partly consistent with Boivin and Ng (2005). According to the authors, there are only negligible differences between the iterated and direct forecasts for inflation variables, but some gains of applying an iterated rather than a direct multi-step forecast scheme to output variables.

Regarding our variables capturing the publication strategy, no clear results appear. While we find an author bias, but no publication bias for output, PUBL_YES and AUTHOR_YES enter significantly in the equation for inflation with a negative and a positive sign, respectively.

4.2.2. Robustness checks: outliers, sampling bias, dependency and qualitative differences

The appropriate weighting of observations is a key issue. We begin this subsection by presenting alternative ways to deal with outliers. We then address problems which may arise due to sampling bias and dependency of observations in a second step and differences in quality in a third step.

Outliers

Our baseline meta-regression (as well as the descriptive analysis) relies on a dataset from which outliers were removed in a particular, rather *ad hoc* manner, as explained in section 3. As a robustness check, we adopt another *ad hoc* outlier removal technique which involves dropping the upper and the lower 5 percentiles of the data, following Knell and Stix (2005b). An alternative way to deal with outliers is to apply robust regression methods to the entire dataset (including those observations which were detected as outliers by our *ad hoc* methods and removed from the sample) (cf. Rousseeuw and Leroy 1987). We adopt a so-called M-estimator which falls into the class of robust regression methods. Instead of minimizing the sum of squared residuals, M-estimators minimize the sum of a function of the residuals. A

²¹ All these studies allow estimation techniques to be disentangled from whether the forecasting equation is estimated restrictedly or unrestrictedly. Schumacher (2006) compares the three factor estimation techniques while estimating forecasting equations unrestrictedly. Forni et al. (2003) and Stock and Watson (2005) compare the FHLR and SW techniques while performing only a restricted and only an unrestricted estimation of the forecasting equation, respectively. Finally, Boivin and Ng (2005) and D’Agostino and Giannone (2006) consider both estimation techniques and perform restricted and unrestricted estimations for both.

number of functions were proposed in the literature which weight down observations with residuals being distant from zero. We use the commonly employed Tukey or biweight weighting function proposed by Beaton and Tukey (1974) where weights gradually decrease with the distance of the residuals from zero and are set to zero from some point on.²² The M-estimation was implemented with iterated reweighted least squares.

Sampling bias and Dependency

There are two problems from using many observations from individual studies. First, observations from the same study (conducted by the same researcher(s)) may not be independent, but may cluster within individual studies. And second, studies which report a large number of results would be given a large weight relative to studies which report only a few results, which is also referred to as sampling bias (cf. Stanley 2001, Weichselbaumer and Winter-Ebmer 2005). These problems could be present in our sample where the number of results used for our analysis ranges from 3 (Grenouilleau 2004) to more than 8,800 (Lin and Tsay 2005) (after removal of outliers and observations with missing values for one or more meta-independent variables). Likewise, dependency and sampling bias may be present because researchers rely on similar or identical datasets from which they estimate the factors. For example, the dataset constructed by SW for the US (or a slightly modified version) is used in a total of 13 studies.

We address these problems in the following manners. First, to eliminate a possible sampling bias and enhance efficiency, we construct a weighting matrix which attributes equal weight to each observation from a particular study $1/n_{g^s}$ where n_{g^s} is the number of observations taken from study $g^s = 1, \dots, 46$. Accordingly, we attribute equal weights to results produced based on equal or similar datasets, namely $1/n_{g^d}$ where n_{g^d} is the number of observations produced with the dataset $g^d = 1, \dots, 31$. Equation (8) is then estimated with weighted least squares (WLS) (see also Weichselbaumer and Winter-Ebmer 2005, Knell and Stix 2005). Second, we apply robust clustering to equation (8) which provides us with robust covariance estimates and thereby adjusts for within-cluster correlation, where each study and each dataset, respectively, represents a cluster.²³

²² The weights equal $[1 - (\eta_j / \kappa_T)^2]^2$ for $|\eta_j| \leq \kappa_T$ and 0 for $|\eta_j| > \kappa_T$, where κ_T , the tuning constant associated with the Tukey or biweight function, is set at 4.685.

²³ Under robust clustering, the estimated variance of the parameters is given by

$\left(\sum_g G_g' G_g \right)^{-1} \left(\sum_g G_g' \eta_g \eta_g' G_g \right) \left(\sum_g G_g' G_g \right)^{-1}$, where $G = [1 \quad M']$ and subscript $g \in \{g^s, g^d\}$ refers to the group (the study or the dataset).

Quality

Many meta-analyses downweight low-quality results and give high-quality results larger weights (cf. Knell and Stix 2005, Weichselbaumer and Winter-Ebmer 2005). Most existing meta-analyses focus on regression coefficient estimates. The underlying studies generally also report some measure of estimation uncertainty such as standard errors, and attribute weights to the observations which are inversely related to this uncertainty. Measuring the quality of observations is, however, difficult in our context (as well as in other contexts: see, for example, Weichselbaumer and Winter-Ebmer 2005). Some of the papers on which our analysis is based do not provide standard errors at all. Others provide standard errors or use diverse formal tests to find out whether some models significantly outperform others.²⁴ The different statistics are, however, barely comparable. Finding a common metric of estimation precision is further complicated by the fact that we often needed to convert results reported in the papers to relative RMSEs, and it is unclear how such a conversion of the statistics translates to the uncertainty surrounding them.

Nevertheless, we try to address the issue of qualitative differences between observations. We follow Weichselbaumer and Winter-Ebmer (2005) and Longhi et al. (2005) and construct weights equal to the inverse of the standard deviation of relative RMSEs within individual studies and fit WLS to equation (8). This presumes that large heterogeneity among relative RMSEs obtained from an individual study goes along with less precise forecasts. Obviously, the drawback of such a weighting scheme is that all observations taken from one study are weighted equally, although some results of that particular study are probably more accurate than others. Also, large heterogeneity across results from an individual study might simply reflect the broad scope of that study and careful consideration of lots of models. Results should therefore be interpreted with caution.

Overall, results are relatively robust across different outlier adjustment methods and weighting schemes, see specifications (3) to (9) in Tables 3 and 4.²⁵ One remark is, however, in order. Dependency seems to be an important issue (cf specifications (7) and (8) in Tables 3 and 4). This is reflected in generally much lower coefficients' *t*-statistics. Some variables turn from significant to insignificant, such as BENCH_RW, T and AUTHOR_YES for both output and inflation. Also, COUNTRY_UK, COUNTRY_OTHER, N and ITDIR_1_STEP no longer enter the equation for output significantly, and BENCH_VAR, PRESEL_1, FACTOR_KM, and PUBL_YES become irrelevant in the equation for inflation. Dependency seems to be present between observations taken from individual studies as well as datasets.

²⁴ They use, for instance, the Diebold and Mariano (1995) tests for non-nested models and the Clark and McCracken (2001) or Giacomini and White (2006) tests for nested models.

²⁵ Interestingly, WLS permits to noticeably reduce the dispersion in the dependent variable, which is reflected in a large adjusted R² of the WLS regressions; see Tables 3 and 4.

5. Conclusion

In this paper, we have taken a meta-analytic approach to assess empirically the relative forecast performance of large dynamic factor models for real economic activity and inflation. This approach has allowed us to systematically summarize findings from existing factor forecast applications. At the same time, we have been able to assess the relevance of a large number of determinants of the factor forecast performance. Most determinants are derived from theory, and potentially affect the precision of factor estimates, the commonality of the target variable and the specification of the forecasting equation.

Our results broadly suggest that, on average, large-scale dynamic factor models have a forecast performance which is slightly superior to that of other models, with inflation forecasts having a slight advantage over output forecasts. This, however, is associated with a greater dispersion of relative RMSEs associated with inflation. Moreover, some of the variables governing the precision of factor estimates and the commonality of the target variable seem to matter for the relative factor forecast performance, whereas variables capturing the specification of the factor forecasting equation turn out to be rather irrelevant.

Let us summarize our main results regarding the determinants of the relative factor forecast performance in some more detail. Among the variables determining the forecast environment (which cannot be influenced by the forecasters), the variable itself (whether output or inflation is predicted and from which country/country group the variable is taken) seems to play a role: Factor models deliver better predictions of US compared to UK variables and of US compared to euro-area output, yet no significant differences were found between US and euro-area inflation. According to our analysis, factor models perform slightly worse than alternative methods which are also designed to handle large datasets. By contrast, they generally outdo small-scale models. Factor models also seem to perform better for output at shorter forecast horizons, while the effect of the horizon on the relative factor forecast performance is insignificant in the case of inflation.

Among the variables governing the forecast design (which can be influenced by the forecaster), the size of the dataset, i.e. the cross-section and time dimensions, positively affects the relative factor forecast performance. Whether forecasters rely on a balanced or an unbalanced panel does not seem to affect the relative factor forecast performance significantly. Likewise (and surprisingly), there is no evidence that pre-selecting of the variables has positively influenced the factor forecasts in the past. By contrast - and this is a very strong result - rolling windows are found to lend themselves better to factor forecasts than recursive windows suggesting that, if structural breaks are presumed to be present in the datasets, it may be advisable to rely on a rolling window. We further find some evidence that quarterly data appear to be better suited to factor models than monthly data. In addition, the technique used to estimate the factors matters: we find that the dynamic approaches of FHLR and KM tend to outperform the static SW approach. This is remarkable given that most

practitioners apply the SW approach, which has the advantage that it is easier to implement. Among the variables describing the specification of the factor forecast equation, we do not find evidence that imposing restrictions implied by the dynamic factor structure in the forecasting equation is harmful, nor do we find significant differences between iterated and direct multi-step forecasts.

We hope that our analysis will help practitioners to improve their forecasts. Our main messages are the following. First, it seems worthwhile exploiting information available in data-rich environments. Second, it is important to carefully specify the forecasting model. Although more demanding in terms of specification, the complex FHLR and KM factor estimation approaches tend to outperform the SW approach, and restricted estimation of the forecasting equation does not seem to be inferior to unrestricted estimation. Third, some potential determinants were found to be surprisingly irrelevant in the empirical analysis, given that researchers have paid relatively much attention to them and wrote many papers about these issues. This refers, for instance, to whether data selection has taken place prior to the factor estimation, to whether or not restrictions implied by the factor structure are imposed in the forecasting equation and whether an unbalanced or a balanced panel is used. Another determinant, by contrast, turns out to be surprisingly important and robust, given that not much attention is paid to it in the existing factor forecast literature: whether the forecast is made based on a rolling or a recursive window. Our fourth and final point is that very recent approaches which were beyond the scope of this paper, such as factor forecast pooling, may be promising given the good performance of factor models and forecast combinations (which form the largest subgroup in BENCH_LARGE, the group of alternative methods suited to handle large datasets) and initial encouraging results, for instance by Koop and Potter (2004) and Stock and Watson (2006).

Table 1: Studies included in the meta-analysis

Paper	# obs	Variable	Country	Factor estim. method
Aguirre, Céspedes (2004)	24	Output, inflation	CHL	SW
Angelini, Henry, Mestre (2001)	4050	Inflation	EA	SW
Artis, Banerjee, Marcellino (2005)	839	Output, inflation	UK	SW
Banerjee, Marcellino (2006)	2470	Output, inflation	US	SW
Banerjee, Marcellino, Masten (2005)	120	Output, inflation	EA	SW
Banerjee, Marcellino, Masten (2006)	2805	Output, inflation	CZ, HU, PL, SK, SI	SW
Boivin, Ng (2005)	240	Output, inflation	US	SW, FHLR
Boivin, Ng (2006)	144	Output, inflation	US	SW
Brisson, Campell, Galbraith (2003)	159	Output, inflation	CAN	SW
Bruneau, de Bandt, Flageollet (2003a)	124	Inflation	EA	SW
Bruneau, de Bandt, Flageollet, Michaux (2003b)	471	Inflation	FRA	SW
Camacho, Sancho (2003)	24	Output, inflation	ESP	SW
Camba-Méndez, Kapetanios (2005)	768	Inflation	EA, DEU, ESP, FRA, ITA, NLD	KM
Cristadoro, Forni, Reichlin, Veronese (2005)	84	Inflation	EA	FHLR
D'Agostino, Giannone (2006)	558	Output, inflation	US	SW, FHLR
Den Reijer (2005)	48	Output	NLD	SW, FHLR
Favero, Ricchi, Tegami (2004)	156	Inflation	ITA	SW
Forni, Hallin, Lippi, Reichlin (2002)	16	Output, inflation	EA	FHLR
Forni, Hallin, Lippi, Reichlin (2003)	96	Output, inflation	EA	SW, FHLR
Gavin, Kliesen (2006)	120	Output, inflation	US	SW
Giacconi, White (2006)	36	Output, inflation	US	SW
Gosselin, Tkacz (2001)	12	Inflation	CAN	SW
Grenouilleau (2004)	3	Output	EA	SW
Inoue, Kilian (2005)	521	Inflation	US	SW
Jeon (2004)	840	Output, inflation	CAN, FRA, DEU, JPN, UK, US	SW
Kapetanios, Labhard, Price (2005)	864	Inflation	UK	SW
Kapetanios (2004)	16	Inflation	UK	KM
Lin, Tsay (2005)	9360	Output	US	SW
Liu (2004)	480	Output, inflation	US	SW
Marcellino, Stock, Watson (2001, 2003)	1767	Output, inflation	EA, AUT, BEL, FIN, FRA, DEU, IRE, ITA, LUX, NLD, PRT,	SW
Matheson (2006)	4928	Output, inflation	NZL	SW
Moser, Rumler, Scharler (2004)	12	Inflation	AUT	SW
Schneider, Spitzer (2004)	6	Output	AUT	FHLR

Table 1: Studies included in the meta-analysis cont.

Paper	# obs	Variable	Country	Factor estim. method
Schumacher (2006)	188	Output	DEU	SW, FHLR, KM
Schumacher, Dreger (2004)	48	Output	DEU	SW
Stavrev (2006)	2778	Inflation	EA	FHLR
Stock, Watson (1998)	410	Output, inflation	US	SW
Stock, Watson (1999)	6184	Inflation	US	SW
Stock, Watson (2002a)	864	Output, inflation	US	SW
Stock, Watson (2002b)	18	Output	US	SW
Stock, Watson (2004)	2028	Output	CAN, FRA, DEU, ITA, JPN, UK, US	SW
Stock, Watson (2005)	336	Output, inflation	US	SW, FHLR
Stock, Watson (2006)	48	Output	US	SW, FHLR
Tatiwa Ferreira, Bierens, Castelar (2005)	25	Output	BRA	SW
Van Nieuwenhuyze (2006)	6	Output	BEL	FHLR
Watson (2003)	20	Output, inflation	US	SW

Notes: # obs refers to the number of observations before outlier removal. We refer to the unpublished version of Forni, Hallin, Lippi, Reichlin (2002) since no forecast performance measures were reported in the published version. Abbreviations are EA: euro area, AUT: Austria, BEL: Belgium, FIN: Finland, FRA: France, DEU: Germany, ITA: Italy, IRE: Ireland, LUX: Luxembourg, NLD: Netherlands, PRT: Portugal, ESP: Spain, NZL: New Zealand, CHL: Chile, BRA: Brazil, CAN: Canada, JPN: Japan, US: United States, UK: United Kingdom, CZ: Czech Republic, HU: Hungary, SI: Slovenia, PL: Poland, SK: Slovakia.

Table 2: Descriptive statistics

	Mean		Median		Std.	Min.	Max.	# obs
Output and Inflation								
Total	0.981	**	0.984	**	0.167	0.525	1.454	38,729
ARIMA	0.998		0.994	**	0.180	0.525	1.454	8,111
RW	0.900	**	0.854	**	0.183	0.528	1.425	623
VAR	0.968	**	0.962	**	0.193	0.525	1.454	2,785
INDIC	0.938	**	0.945	**	0.121	0.526	1.454	13,413
LARGE	1.020	**	1.011	**	0.180	0.526	1.454	13,797
US	0.969	**	0.972	**	0.126	0.525	1.454	20,012
UK	1.031	**	1.023	**	0.149	0.528	1.453	2,062
EA	0.956	**	0.975	**	0.207	0.525	1.454	9,791
OTHERS	1.037	**	1.039	**	0.194	0.526	1.454	6,864
Output								
Total	0.988	**	0.988	**	0.142	0.525	1.454	18,910
ARIMA	0.967	**	0.970	**	0.171	0.525	1.454	3,840
RW	0.879	**	0.843	**	0.158	0.528	1.245	274
VAR	0.945	**	0.927	**	0.178	0.528	1.452	1,311
INDIC	0.939	**	0.929	**	0.181	0.526	1.454	1,919
LARGE	1.010	**	1.005	**	0.110	0.526	1.454	11,566
US	0.978	**	0.980	**	0.123	0.525	1.449	11,968
UK	0.990	(*)	1.003		0.137	0.528	1.453	741
EA	1.002		1.010		0.146	0.535	1.449	2,600
OTHERS	1.008	*	1.005	*	0.189	0.526	1.454	3,601
Inflation								
Total	0.975	**	0.980	**	0.187	0.525	1.454	19,819
ARIMA	1.026	**	1.020	**	0.183	0.530	1.452	4,271
RW	0.916	**	0.860	**	0.199	0.530	1.425	349
VAR	0.989	(*)	0.995		0.204	0.525	1.454	1,474
INDIC	0.938	**	0.948	**	0.180	0.526	1.454	11,494
LARGE	1.069	**	1.057	**	0.157	0.531	1.454	2,231
US	0.955	**	0.959	**	0.130	0.526	1.454	8,044
UK	1.055	**	1.040	**	0.151	0.637	1.451	1,321
EA	0.940	**	0.953	**	0.223	0.525	1.454	7,191
OTHERS	1.070	**	1.079	**	0.193	0.526	1.454	3,263

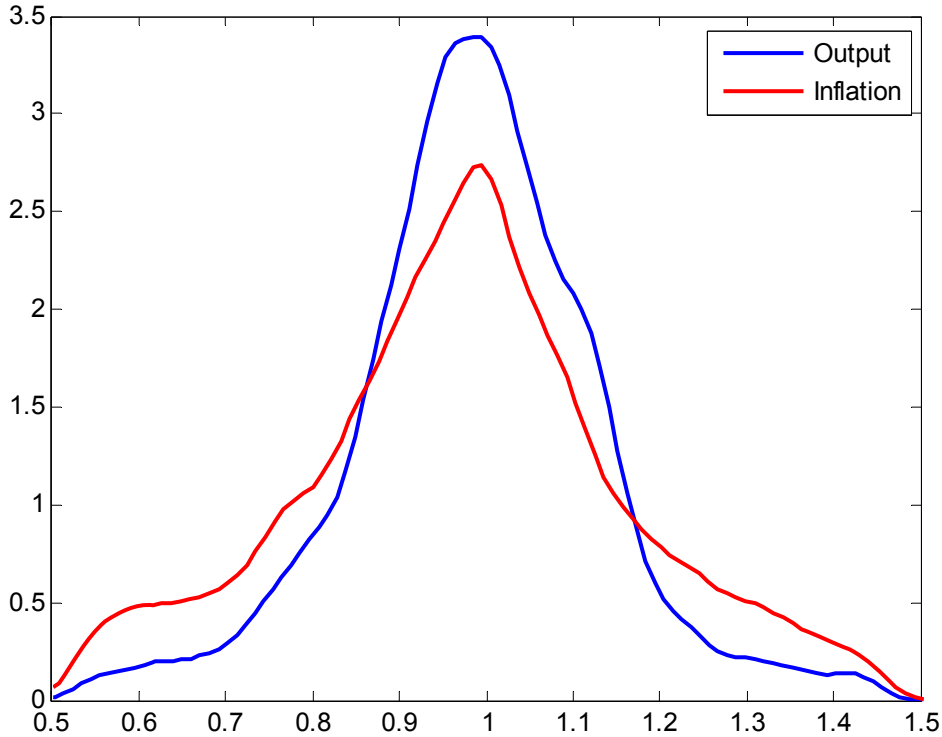
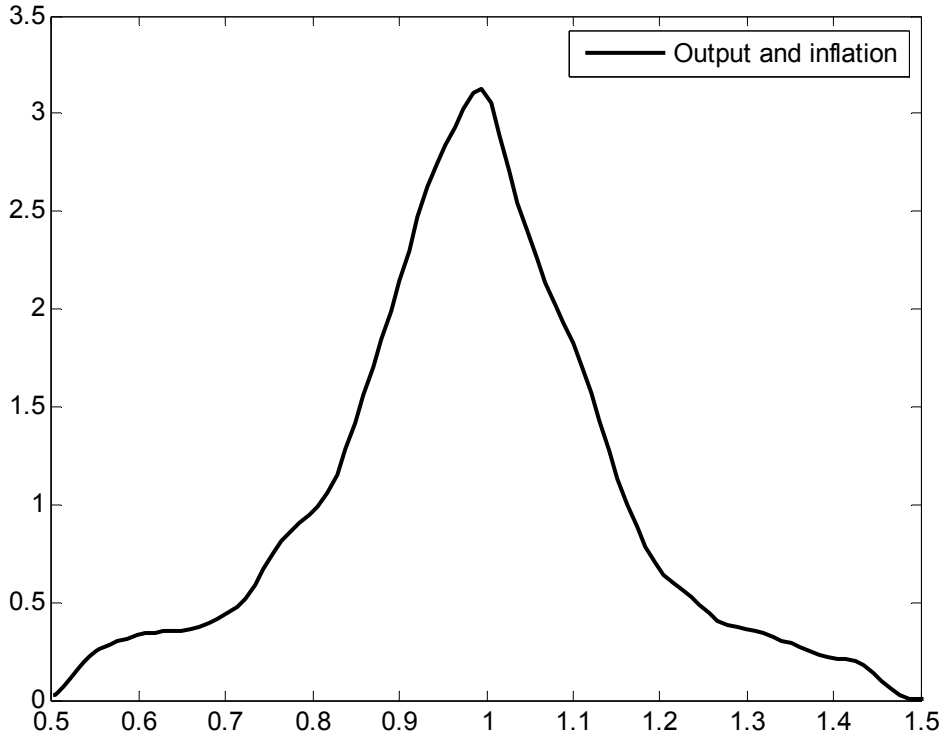
Notes: **/(*) indicates that values are significantly different from 1 at the 1/5/10 percent level. This is tested using a t -test for the means and a Wilcoxon-sign rank test for the medians.

Table 4: Meta-regression results for inflation

	Baseline			Outliers			Sampling bias			Dependency			Quality				
	(1) coef	(2) t-stat	(2) coef	(3) coef	(3) t-stat	(4) coef	(4) t-stat	(5) coef	(5) t-stat	(6) coef	(6) t-stat	(7) coef	(7) t-stat	(8) coef	(8) t-stat	(9) coef	(9) t-stat
CONSTANT	1.06	132.27	1.06	1.10	114.11	1.19	63.84	1.04	112.96	1.03	123.42	1.06	18.69	1.06	17.53	1.07	116.85
COUNTRY_UK	0.05	5.76	0.04	0.03	3.46	0.01	0.32	0.04	3.89	0.03	3.70	0.05	2.20	0.05	2.06	0.06	6.17
COUNTRY_EA	0.00	0.56	-	0.03	3.40	0.02	1.39	-0.01	-0.72	-0.01	-1.41	0.00	0.09	0.00	0.09	0.01	0.73
COUNTRY_OTHER	0.07	3.86	0.06	0.08	3.90	0.09	2.35	0.05	2.09	0.02	1.03	0.07	3.33	0.07	3.16	0.09	4.37
BENCH_RW	-0.07	-6.03	-0.07	-0.08	-5.81	-0.10	-3.92	-0.07	-5.56	-0.07	-5.43	-0.07	-1.77	-0.07	-1.73	-0.08	-5.40
BENCH_VAR	-0.02	-2.11	-	-0.03	-3.20	-0.06	-4.22	0.00	0.08	0.00	0.12	-0.02	-1.41	-0.02	-1.34	-0.02	-2.98
BENCH_INDIC	-0.08	-15.64	-0.08	-0.10	-15.61	-0.13	-10.42	-0.07	-10.39	-0.07	-11.92	-0.08	-1.96	-0.08	-1.84	-0.10	-16.36
BENCH_LARGE	0.04	7.13	0.04	0.03	4.71	0.01	0.68	0.04	5.71	0.03	5.16	0.04	1.76	0.04	1.70	0.04	6.89
HORIZON (demeaned)	-0.00	-0.93	-	0.01	5.82	0.01	6.30	-0.00	-1.43	-0.00	-1.23	-0.00	-0.21	-0.00	-0.19	-0.00	-0.78
log(N) (demeaned)	-0.01	-4.99	-0.01	-0.02	-5.61	-0.01	-2.11	-0.01	-3.44	-0.01	-4.08	-0.01	-2.43	-0.01	-2.71	-0.01	-4.70
log(T) (demeaned)	-0.04	-5.12	-0.05	-0.07	-7.45	-0.12	-7.62	-0.04	-3.60	-0.03	-3.47	-0.04	-1.37	-0.04	-1.27	-0.05	-4.85
BAL_NO	0.01	1.08	-	0.00	0.74	-0.00	-0.21	0.00	0.34	-0.00	-0.29	0.01	0.56	0.01	0.50	0.01	1.00
PRE_SEL_1	0.06	8.29	0.06	0.06	7.04	0.06	4.04	0.07	7.16	0.07	7.71	0.06	1.73	0.06	1.64	0.06	7.59
PRE_SEL_2	0.04	10.47	0.04	0.03	7.37	0.03	3.49	0.03	8.42	0.04	10.17	0.04	3.64	0.04	3.68	0.03	8.28
ROLREC_ROL	-0.16	-6.82	-0.16	-0.20	-6.93	-0.31	-8.33	-0.15	-4.89	-0.15	-5.66	-0.16	-2.25	-0.16	-2.10	-0.16	-5.85
FREQ_Q	-0.10	-8.67	-0.11	-0.15	-10.63	-0.23	-10.59	-0.08	-5.98	-0.07	-5.19	-0.10	-2.14	-0.10	-1.94	-0.12	-8.79
FACTOR_FHLR	-0.12	-10.39	-0.12	-0.15	-10.87	-0.32	-14.87	-0.12	-9.50	-0.09	-8.28	-0.12	-2.08	-0.12	-1.98	-0.13	-9.74
FACTOR_KM	-0.03	-1.98	-0.05	-0.08	-4.62	-0.14	-3.35	-0.03	-2.00	-0.03	-2.10	-0.03	-0.48	-0.03	-0.45	-0.03	-1.93
RESTR_YES	-0.02	-1.57	-	-0.02	-1.61	0.08	3.05	0.01	0.80	-0.00	-0.06	-0.02	-0.56	-0.02	-0.54	-0.03	-2.18
ITDIR_IT	0.00	0.30	-	-0.03	-3.74	-0.07	-5.38	0.01	0.77	0.01	1.45	0.00	0.10	0.00	0.09	0.00	0.48
ITDIR_I_STEP	-0.04	-6.47	-0.04	-0.04	-5.77	-0.04	-2.89	-0.03	-4.94	-0.02	-3.71	-0.04	-4.87	-0.04	-3.55	-0.04	-6.31
PUBL_YES	-0.04	-4.57	-0.04	-0.03	-3.00	-0.02	-0.88	-0.04	-3.32	-0.03	-2.60	-0.04	-1.41	-0.04	-1.42	-0.05	-4.69
AUTHOR_YES	0.04	6.89	0.04	0.04	4.84	0.02	1.24	0.04	5.06	0.04	5.30	0.04	1.01	0.04	0.99	0.06	7.75
# obs	17,294			18,312		20,081		17,294		17,294		17,294		17,294		17,294	
R ² adj	0.167		0.167	0.170		0.090		0.874		0.877		0.167		0.167		0.520	
R ² adj*	-		-	-		-		0.163		0.158		-		-		0.164	

Notes: (1) *Ad hoc* outlier adjustment (OLS with White corrected stdev.), unweighted, (2) as (1), but after successive removal of insignificant variables, (3) *ad hoc* outlier adjustment where the upper and lower percentiles of the data were removed, (4) robust estimation, Tukey or biweight function, (5) *ad hoc* outlier adjustment, WLS (equal weights for each obs. from each study and for each study), (6) *ad hoc* outlier adjustment, WLS (equal weights for each obs. from each dataset and for each dataset), (7) *ad hoc* outlier adjustment, robust clustering where each study represents a cluster, (8) *ad hoc* outlier adjustment, robust clustering where each study represents a cluster, (9) *ad hoc* outlier adjustment, WLS (weights according to standard deviation of residuals within each study). R²adj* refers to the adjusted R² where the WLS coefficient estimates are applied to the unweighted data.

Figure 1: Smoothed histograms of relative RMSEs



Note: after outlier adjustment.

References

- Altissimo, F., A. Bassanetti, R. Cristadoro, M. Forni, M. Hallin, M. Lippi, L. Reichlin (2001), “EuroCOIN: a real time coincident indicator of the euro area business cycle”, CEPR Working Paper 3108.
- Aguirre, A., L. F. Céspedes (2004), “Uso de análisis factorial dinámico para proyecciones macroeconómicas”, Working Papers Central Bank of Chile 274.
- Angelini, E., J. Henry, R. Mestre (2001), “Diffusion index-based inflation forecasts for the euro area”, ECB Working Paper 61.
- Artis, M. J., A. Banerjee, M. Marcellino (2005), “Factor forecasts for the UK”, *Journal of Forecasting*, 24(4), 279-298.
- Bai, J., S. Ng (2002), “Determining the number of factors in approximate factor models”, *Econometrica*, 70(1), 191-221.
- Bai, J., S. Ng (2005), “Determining the number of primitive shocks in factor models”, *Journal of Business and Economic Statistics*, forthcoming.
- Banerjee, A., M. Marcellino (2006), “Are there any reliable leading indicators for US inflation and GDP growth?” *International Journal of Forecasting*, 22, 137-151.
- Banerjee, A., M. Marcellino, I. Masten (2005), “Leading indicators for euro-area inflation and GDP growth.”, *Oxford Bulletin of Economics and Statistics*, 67, 785-814.
- Banerjee, A., M. Marcellino, I. Masten (2006), “Forecasting macroeconomic variables for the new member states” in: Artis, M.J., A. Banerjee, M. Marcellino (eds.), *The central and eastern European countries and the European Union*, Cambridge University Press, Cambridge, Chapter 4, 108-134.
- Beaton, A.E., J.W. Tukey (1974), “The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data”, *Technometrics*, 16(2), 147-185.
- Boivin, J., S. Ng (2005), “Understanding and comparing factor-based forecasts”, *International Journal of Central Banking*, 1, 117-151.
- Boivin, J., S. Ng (2006), “Are more data always better for factor analysis”, *Journal of Econometrics*, 132, 169-194.
- Breitung, J., S. Eickmeier (2006), “Dynamic factor models”, in: O. Hübler and J. Frohn (Eds.), *Modern econometric analysis*, Chapter 3, Springer 2006.
- Brisson, M., B. Campbell, J.W. Galbraith (2003), “Forecasting some low-predictability time series using diffusion indices”, *Journal of Forecasting*, 22, 515-531.

- Bruneau, C., O. de Bandt, A. Flageollet (2003a), “Forecasting inflation in the euro area”, Banque de France NER 102.
- Bruneau, C., O. de Bandt, A. Flageollet, E. Michaux (2003b), “Forecasting inflation using economic indicators: the case of France”, Banque de France NER 101.
- Burns, A. F., W.C. Mitchell (1946), “Measuring business cycles”, New York, NBER.
- Camacho, M., I. Sancho (2003), “Spanish diffusion indexes”, *Spanish Economic Review*, 5, 173-203.
- Camba-Méndez, G., G. Kapetanios (2005), “Forecasting euro area inflation using dynamic factor measures of underlying inflation”, *Journal of Forecasting*, 25, 491-503.
- Clark, T., M. McCracken (2001), “Tests of equal forecast accuracy and encompassing for nested models”, *Journal of Econometrics*, 105, 85-110.
- Cristadoro, R., M. Forni, L. Reichlin, G. Veronese (2005), “A core inflation indicator for the euro area”, *Journal of Money, Credit and Banking*, 37(3), 539-560.
- D’Agostino, A., D. Giannone (2006), “Comparing alternative predictors based on large-panel factor models”, ECB Working Paper 680.
- Den Reijer, A. (2005), “Forecasting Dutch GDP using large scale factor models”, DNB Working Paper 28.
- Diebold, F., R. Mariano (1995), “Comparing predictive accuracy”, *Journal of Business & Economic Statistics*, 13, 253-263.
- Doucouliafos, C. (2005), “Publication bias in the economic freedom and economic growth literature”, *Journal of Economic Surveys*, 19(3), 367-387.
- Favero, C., O. Ricchi, C. Tegami (2004), “Forecasting Italian inflation with large datasets and many models”, IGIER Working Paper 269.
- Fidrmuc, J., I. Korhonen (2006), “Meta-analysis of the business cycle correlation between the euro area and the CEECs”, *Journal of Comparative Economics*, 34(3), 518-537.
- Forni, M., M. Hallin, M. Lippi, L. Reichlin (2005), “The generalized dynamic factor model: one-sided estimation and forecasting”, *Journal of the American Statistical Association*, 100(471), 830-840 (11).
- Forni, M., M. Hallin, M. Lippi, L. Reichlin (2002), “The generalized dynamic factor model: one-sided estimation and forecasting”, CEPR Discussion Paper 3432.
- Forni, M., M. Hallin, M. Lippi, L. Reichlin (2000), “The generalized dynamic factor model: identification and estimation”, *The Review of Economic and Statistics*, 82, 540-554.

- Forni, M., M. Hallin, M. Lippi, L. Reichlin (2003), “Do financial variables help forecasting inflation and real activity in the euro area? “, *Journal of Monetary Economics*, 50, 1243-1255.
- Forni, M., M. Hallin, M. Lippi, L. Reichlin (2005), “The generalized dynamic factor model: one-sided estimation and forecasting”, *Journal of the American Statistical Association*, 100, 830-840.
- Gavin, W.T., K.L. Kliesen (2006), “Forecasting inflation and output: comparing data-rich models with simple rules”, Federal Reserve Bank of St. Louis, Working Paper 54A.
- Giacomini, R., H. White (2006), “Tests of conditional predictive ability”, *Econometrica*, 74(6), 1545–1578.
- Giannone, D., L. Reichlin, L. Sala (2002), ”Tracking Greenspan: systematic and unsystematic monetary policy revisited”, CEPR Discussion Paper 3550.
- Giannone, D., L. Reichlin, L. Sala (2004), “Monetary policy in real time”, in: Gertler, M., K. Rogoff (eds.), *NBER Macroeconomics Annual*, MIT Press, 161-200.
- Gosselin, M.-A., G. Tkacz (2001), “Evaluating factor models: an application to forecasting inflation in Canada”, Bank of Canada Working Paper 18.
- Grenouilleau, D. (2004), “A sorted leading indicators dynamic (SLID) factor model for short-run euro-area GDP forecasting”, *European Commission Economic Papers* 219.
- Huber, P.J. (1964), “Robust estimation of a location parameter“, *Annals of Mathematical Statistics*, 35(1), 73-101.
- Inoue, A., L. Kilian (2005), “How useful is bagging in forecasting economic time series? A case study of US CPI inflation”, CEPR Discussion Paper 5304.
- Jeon, Y. (2004), “Combined forecasts, hidden common factors and international linkages”, mimeo, <http://www.nd.edu/~meg/MEG2004/Jeon-Yongil.pdf>.
- Kapetanios, G., V. Labhard, S. Price (2005), “Forecasting using Bayesian and information theoretic model averaging: an application to UK inflation”, *Journal of Economic and Business Statistics*, forthcoming.
- Kapetanios, G. (2004), “A note on modelling core inflation for the UK using a new dynamic factor estimation method and a large disaggregated price index dataset”, *Economics Letters*, 85, 63-69.
- Kapetanios, G., M. Marcellino (2004), “A parametric estimation method for dynamic factor models of large dimensions”, Queen Mary University of London Working Paper 489 revised February 2004.
- Knell, M., H. Stix (2005a), “The income elasticity of money demand: A meta-analysis of empirical results”, *Journal of Economic Surveys*, 19(3), 213-233.

- Knell, M., H. Stix (2005b), “How robust are money demand estimations? A meta-analytic summary of findings about income elasticities”, *Kredit und Kapital*, 38 (4), 515-540.
- Knell, M., H. Stix (2006), “Three decades of money demand studies: differences and similarities”, *Applied Economics*, 38, 805-818.
- Koop, G., S. Potter (2004), Forecasting in dynamic factor models using Bayesian model averaging”, *Econometrics Journal*, 7, 550-565.
- Lin, J.-L., R.S. Tsay (2005), “Comparison of forecasting methods with many predictors”, mimeo, <http://www.atl-res.com/finance/LIN.pdf>.
- Lipsey, M.W., D. B. Wilson (2000), “Practical meta-analysis”, SAGE Publications Ltd.
- Liu, D. (2004), “The effects of monetary policy and macroeconomic forecasting”, mimeo, <http://econweb.tamu.edu/dliu/The%20Effects%20of%20Monetary%20and%20Forecasting%20oct31.pdf>.
- Longhi, S., P. Nijkamp, J. Poot (2005), “ A meta-analytic assessment of the effect of immigration on wages”, *Journal of Economic Surveys*, 19(3), 451-477.
- Marcellino, M., J. Stock, M. Watson (2001), “Macroeconomic forecasting in the euro area: country-specific versus area-wide information”, IGIER Working Paper 201.
- Marcellino, M., J. Stock, M. Watson (2003), “Macroeconomic forecasting in the euro area: country-specific versus area-wide information”, *European Economic Review*, 47, 1-18.
- Marcellino, M., J. Stock, M. Watson (2005), “A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series”, *Journal of Econometrics*, 135, 499–526.
- Matheson, T. D. (2006), “Factor model forecasts for New Zealand”, *International Journal of Central Banking*, 2, 169-237.
- Moser, G., F. Rumler, J. Scharler (2004), “Forecasting Austrian inflation”, OeNB Working Paper 91.
- Reichlin, L. (2003), “Factor models in large cross sections of time series”, in: Dewatripont, M., L. Hansen, S. Turnovsky (eds), *Advances in Econometrics, Theory and Applications*, Eight World Congress of the Econometric Society, Vol. III.
- Rousseeuw, P.J., A.M. Leroy (1987), “Robust regression and outlier detection “, Jon Wiley & Sons Inc., New York.
- Schneider, M., M. Spitzer (2004), “Forecasting Austrian GDP using the generalized dynamic factor model”, OeNB Working Paper 89.
- Schumacher, C. (2006), “Forecasting German GDP using alternative factor models based on large dataset”, *Journal of Forecasting*, forthcoming.

- Schumacher, C. , J. Breitung (2006), “Real-time forecasting of GDP based on a large factor model with monthly and quarterly data”, Bundesbank Discussion Paper, Series 1, 33/2006.
- Schumacher, C., C. Dreger (2004), “Estimating large-scale factor models for economic activity in Germany: Do they outperform simpler models?“, *Jahrbücher für Nationalökonomie und Statistik*, 224, 731-750.
- Stanley, T.S. (2001), “Wheat from chaff: meta-analysis as quantitative literature review”, *Journal of Economic Perspectives*, 15(3), 131-150.
- Stavrev, E. (2006), “Measures of underlying inflation in the euro area: assessment and role for informing monetary policy”, IMF Working Paper 197.
- Stock, J., M. Watson (1998), “Diffusion indexes”, NBER Working Paper 6702.
- Stock, J., M. Watson (1999), “Forecasting inflation”, *Journal of Monetary Economics*, 44, 293-335.
- Stock, J., M. Watson (2002a), “Macroeconomic forecasting using diffusion indexes”, *Journal of Business and Economic Statistics*, 20,147-162.
- Stock, J., M. Watson (2002b), “Forecasting using principal components from a large number of predictors”, *Journal of the American Statistical Association*, 97, 1167-1179.
- Stock, J., M. Watson (2004), “Combination forecasts of output growth in a seven country dataset”, *Journal of Forecasting*, 23(6), 405-430.
- Stock, J., M. Watson (2005), “An empirical comparison of methods for forecasting using many predictors”, mimeo, http://ksghome.harvard.edu/~JStock/pdf/beb_4.pdf.
- Stock, J., M. Watson (2006), “Forecasting with many predictors”, in: Elliott, G., C.W.J. Granger, A. Timmermann (eds.), *Handbook of economic forecasting*, Vol. 1, 515-554.
- Tatiwa Ferreira, R., H. Bierens, I. Castelar (2005), “Forecasting quarterly Brazilian GDP growth rate with linear and nonlinear diffusion index models”, *EconomiA, Selecta, Brasilia (DF)*, 6(3), 261-292.
- Uhlig, H. (2004), “What moves GNP ?”, mimeo, presented at the Econometric Society North American Winter Meeting.
- Van Nieuwenhuyze, C. (2006), “A generalized dynamic model for the Belgian Economy - useful business cycle indicators and GDP growth forecasts”, National Bank of Belgium Working Paper 80.
- Watson, M. (2003), “Macroeconomic forecasting using many predictors”, in: Dewatripont, M., L. Hansen, S. Turnovsky (eds.), *Advances in Econometrics, Theory and Applications*, Eight World Congress of the Econometric Society, Vol. III, 87-115.

Weichselbaumer, D., R. Winter-Ebmer (2005), "A meta-analysis of the international wage gender gap", *Journal of Economic Surveys*, 19(3), 479-511.

The following Discussion Papers have been published since 2005:

Series 1: Economic Studies

1	2005	Financial constraints and capacity adjustment in the United Kingdom – Evidence from a large panel of survey data	Ulf von Kalckreuth Emma Murphy
2	2005	Common stationary and non-stationary factors in the euro area analyzed in a large-scale factor model	Sandra Eickmeier
3	2005	Financial intermediaries, markets, and growth	F. Fecht, K. Huang, A. Martin
4	2005	The New Keynesian Phillips Curve in Europe: does it fit or does it fail?	Peter Tillmann
5	2005	Taxes and the financial structure of German inward FDI	Fred Ramb A. J. Weichenrieder
6	2005	International diversification at home and abroad	Fang Cai Francis E. Warnock
7	2005	Multinational enterprises, international trade, and productivity growth: Firm-level evidence from the United States	Wolfgang Keller Steven R. Yeaple
8	2005	Location choice and employment decisions: a comparison of German and Swedish multinationals	S. O. Becker, K. Ekholm, R. Jäckle, M.-A. Muendler
9	2005	Business cycles and FDI: evidence from German sectoral data	Claudia M. Buch Alexander Lipponer
10	2005	Multinational firms, exclusivity, and the degree of backward linkages	Ping Lin Kamal Saggi

11	2005	Firm-level evidence on international stock market comovement	Robin Brooks Marco Del Negro
12	2005	The determinants of intra-firm trade: in search for export-import magnification effects	Peter Egger Michael Pfaffermayr
13	2005	Foreign direct investment, spillovers and absorptive capacity: evidence from quantile regressions	Sourafel Girma Holger Görg
14	2005	Learning on the quick and cheap: gains from trade through imported expertise	James R. Markusen Thomas F. Rutherford
15	2005	Discriminatory auctions with seller discretion: evidence from German treasury auctions	Jörg Rocholl
16	2005	Consumption, wealth and business cycles: why is Germany different?	B. Hamburg, M. Hoffmann, J. Keller
17	2005	Tax incentives and the location of FDI: evidence from a panel of German multinationals	Thiess Buettner Martin Ruf
18	2005	Monetary Disequilibria and the Euro/Dollar Exchange Rate	Dieter Nautz Karsten Ruth
19	2005	Berechnung trendbereinigter Indikatoren für Deutschland mit Hilfe von Filterverfahren	Stefan Stamford
20	2005	How synchronized are central and east European economies with the euro area? Evidence from a structural factor model	Sandra Eickmeier Jörg Breitung
21	2005	Asymptotic distribution of linear unbiased estimators in the presence of heavy-tailed stochastic regressors and residuals	J.-R. Kurz-Kim S.T. Rachev G. Samorodnitsky

22	2005	The Role of Contracting Schemes for the Welfare Costs of Nominal Rigidities over the Business Cycle	Matthias Paustian
23	2005	The cross-sectional dynamics of German business cycles: a bird's eye view	J. Döpke, M. Funke S. Holly, S. Weber
24	2005	Forecasting German GDP using alternative factor models based on large datasets	Christian Schumacher
25	2005	Time-dependent or state-dependent price setting? – micro-evidence from German metal-working industries –	Harald Stahl
26	2005	Money demand and macroeconomic uncertainty	Claus Greiber Wolfgang Lemke
27	2005	In search of distress risk	J. Y. Campbell, J. Hilscher, J. Szilagyi
28	2005	Recursive robust estimation and control without commitment	Lars Peter Hansen Thomas J. Sargent
29	2005	Asset pricing implications of Pareto optimality with private information	N. R. Kocherlakota Luigi Pistaferri
30	2005	Ultra high frequency volatility estimation with dependent microstructure noise	Y. Aït-Sahalia, P. A. Mykland, L. Zhang
31	2005	Umstellung der deutschen VGR auf Vorjahrespreisbasis – Konzept und Konsequenzen für die aktuelle Wirtschaftsanalyse sowie die ökonomische Modellierung	Karl-Heinz Tödter

32	2005	Determinants of current account developments in the central and east European EU member states – consequences for the enlargement of the euro area	Sabine Herrmann Axel Jochem
33	2005	An estimated DSGE model for the German economy within the euro area	Ernest Pytlarczyk
34	2005	Rational inattention: a research agenda	Christopher A. Sims
35	2005	Monetary policy with model uncertainty: distribution forecast targeting	Lars E.O. Svensson Noah Williams
36	2005	Comparing the value relevance of R&D reporting in Germany: standard and selection effects	Fred Ramb Markus Reitzig
37	2005	European inflation expectations dynamics	J. Döpke, J. Dovern U. Fritsche, J. Slacalek
38	2005	Dynamic factor models	Sandra Eickmeier Jörg Breitung
39	2005	Short-run and long-run comovement of GDP and some expenditure aggregates in Germany, France and Italy	Thomas A. Knetsch
40	2005	A “wreckers theory” of financial distress	Ulf von Kalckreuth
41	2005	Trade balances of the central and east European EU member states and the role of foreign direct investment	Sabine Herrmann Axel Jochem
42	2005	Unit roots and cointegration in panels	Jörg Breitung M. Hashem Pesaran
43	2005	Price setting in German manufacturing: new evidence from new survey data	Harald Stahl

1	2006	The dynamic relationship between the Euro overnight rate, the ECB's policy rate and the term spread	Dieter Nautz Christian J. Offermanns
2	2006	Sticky prices in the euro area: a summary of new micro evidence	Álvarez, Dhyne, Hoeberichts Kwapil, Le Bihan, Lünnemann Martins, Sabbatini, Stahl Vermeulen, Vilmunen
3	2006	Going multinational: What are the effects on home market performance?	Robert Jäckle
4	2006	Exports versus FDI in German manufacturing: firm performance and participation in international markets	Jens Matthias Arnold Katrin Hussinger
5	2006	A disaggregated framework for the analysis of structural developments in public finances	Kremer, Braz, Brosens Langenus, Momigliano Spolander
6	2006	Bond pricing when the short term interest rate follows a threshold process	Wolfgang Lemke Theofanis Archontakis
7	2006	Has the impact of key determinants of German exports changed? Results from estimations of Germany's intra euro-area and extra euro-area exports	Kerstin Stahn
8	2006	The coordination channel of foreign exchange intervention: a nonlinear microstructural analysis	Stefan Reitz Mark P. Taylor
9	2006	Capital, labour and productivity: What role do they play in the potential GDP weakness of France, Germany and Italy?	Antonio Bassanetti Jörg Döpke, Roberto Torrini Roberta Zizza
10	2006	Real-time macroeconomic data and ex ante predictability of stock returns	J. Döpke, D. Hartmann C. Pierdzioch

11	2006	The role of real wage rigidity and labor market frictions for unemployment and inflation dynamics	Kai Christoffel Tobias Linzert
12	2006	Forecasting the price of crude oil via convenience yield predictions	Thomas A. Knetsch
13	2006	Foreign direct investment in the enlarged EU: do taxes matter and to what extent?	Guntram B. Wolff
14	2006	Inflation and relative price variability in the euro area: evidence from a panel threshold model	Dieter Nautz Juliane Scharff
15	2006	Internalization and internationalization under competing real options	Jan Hendrik Fisch
16	2006	Consumer price adjustment under the microscope: Germany in a period of low inflation	Johannes Hoffmann Jeong-Ryeol Kurz-Kim
17	2006	Identifying the role of labor markets for monetary policy in an estimated DSGE model	Kai Christoffel Keith Küster Tobias Linzert
18	2006	Do monetary indicators (still) predict euro area inflation?	Boris Hofmann
19	2006	Fool the markets? Creative accounting, fiscal transparency and sovereign risk premia	Kerstin Bernoth Guntram B. Wolff
20	2006	How would formula apportionment in the EU affect the distribution and the size of the corporate tax base? An analysis based on German multinationals	Clemens Fuest Thomas Hemmelgarn Fred Ramb

21	2006	Monetary and fiscal policy interactions in a New Keynesian model with capital accumulation and non-Ricardian consumers	Campbell Leith Leopold von Thadden
22	2006	Real-time forecasting and political stock market anomalies: evidence for the U.S.	Martin Bohl, Jörg Döpke Christian Pierdzioch
23	2006	A reappraisal of the evidence on PPP: a systematic investigation into MA roots in panel unit root tests and their implications	Christoph Fischer Daniel Porath
24	2006	Margins of multinational labor substitution	Sascha O. Becker Marc-Andreas Münder
25	2006	Forecasting with panel data	Badi H. Baltagi
26	2006	Do actions speak louder than words? Household expectations of inflation based on micro consumption data	Atsushi Inoue Lutz Kilian Fatma Burcu Kiraz
27	2006	Learning, structural instability and present value calculations	H. Pesaran, D. Pettenuzzo A. Timmermann
28	2006	Empirical Bayesian density forecasting in Iowa and shrinkage for the Monte Carlo era	Kurt F. Lewis Charles H. Whiteman
29	2006	The within-distribution business cycle dynamics of German firms	Jörg Döpke Sebastian Weber
30	2006	Dependence on external finance: an inherent industry characteristic?	George M. von Furstenberg Ulf von Kalckreuth
31	2006	Comovements and heterogeneity in the euro area analyzed in a non-stationary dynamic factor model	Sandra Eickmeier

32	2006	Forecasting using a large number of predictors: is Bayesian regression a valid alternative to principal components?	Christine De Mol Domenico Giannone Lucrezia Reichlin
33	2006	Real-time forecasting of GDP based on a large factor model with monthly and quarterly data	Christian Schumacher Jörg Breitung
34	2006	Macroeconomic fluctuations and bank lending: evidence for Germany and the euro area	S. Eickmeier B. Hofmann, A. Worms
35	2006	Fiscal institutions, fiscal policy and sovereign risk premia	Mark Hallerberg Guntram B. Wolff
36	2006	Political risk and export promotion: evidence from Germany	C. Moser T. Nestmann, M. Wedow
37	2006	Has the export pricing behaviour of German enterprises changed? Empirical evidence from German sectoral export prices	Kerstin Stahn
38	2006	How to treat benchmark revisions? The case of German production and orders statistics	Thomas A. Knetsch Hans-Eggert Reimers
39	2006	How strong is the impact of exports and other demand components on German import demand? Evidence from euro-area and non-euro-area imports	Claudia Stirböck
40	2006	Does trade openness increase firm-level volatility?	C. M. Buch, J. Döpke H. Strotmann
41	2006	The macroeconomic effects of exogenous fiscal policy shocks in Germany: a disaggregated SVAR analysis	Kirsten H. Heppke-Falk Jörn Tenhofen Guntram B. Wolff

42	2006	How good are dynamic factor models at forecasting output and inflation? A meta-analytic approach	Sandra Eickmeier Christina Ziegler
----	------	--	---------------------------------------

Series 2: Banking and Financial Studies

1	2005	Measurement matters – Input price proxies and bank efficiency in Germany	Michael Koetter
2	2005	The supervisor’s portfolio: the market price risk of German banks from 2001 to 2003 – Analysis and models for risk aggregation	Christoph Memmel Carsten Wehn
3	2005	Do banks diversify loan portfolios? A tentative answer based on individual bank loan portfolios	Andreas Kamp Andreas Pfingsten Daniel Porath
4	2005	Banks, markets, and efficiency	F. Fecht, A. Martin
5	2005	The forecast ability of risk-neutral densities of foreign exchange	Ben Craig Joachim Keller
6	2005	Cyclical implications of minimum capital requirements	Frank Heid
7	2005	Banks’ regulatory capital buffer and the business cycle: evidence for German savings and cooperative banks	Stéphanie Stolz Michael Wedow
8	2005	German bank lending to industrial and non-industrial countries: driven by fundamentals or different treatment?	Thorsten Nestmann
9	2005	Accounting for distress in bank mergers	M. Koetter, J. Bos, F. Heid C. Kool, J. Kolari, D. Porath
10	2005	The eurosystem money market auctions: a banking perspective	Nikolaus Bartzsch Ben Craig, Falko Fecht
11	2005	Financial integration and systemic risk	Falko Fecht Hans Peter Grüner

12	2005	Evaluating the German bank merger wave	Michael Koetter
13	2005	Incorporating prediction and estimation risk in point-in-time credit portfolio models	A. Hamerle, M. Knapp, T. Liebig, N. Wildenauer
14	2005	Time series properties of a rating system based on financial ratios	U. Krüger, M. Stötzel, S. Trück
15	2005	Inefficient or just different? Effects of heterogeneity on bank efficiency scores	J. Bos, F. Heid, M. Koetter, J. Kolatri, C. Kool
01	2006	Forecasting stock market volatility with macroeconomic variables in real time	J. Döpke, D. Hartmann C. Pierdzioch
02	2006	Finance and growth in a bank-based economy: is it quantity or quality that matters?	Michael Koetter Michael Wedow
03	2006	Measuring business sector concentration by an infection model	Klaus Düllmann
04	2006	Heterogeneity in lending and sectoral growth: evidence from German bank-level data	Claudia M. Buch Andrea Schertler Natalja von Westernhagen
05	2006	Does diversification improve the performance of German banks? Evidence from individual bank loan portfolios	Evelyn Hayden Daniel Porath Natalja von Westernhagen
06	2006	Banks' regulatory buffers, liquidity networks and monetary policy transmission	Christian Merkl Stéphanie Stolz
07	2006	Empirical risk analysis of pension insurance – the case of Germany	W. Gerke, F. Mager T. Reinschmidt C. Schmieder

08	2006	The stability of efficiency rankings when risk-preferences and objectives are different	Michael Koetter
09	2006	Sector concentration in loan portfolios and economic capital	Klaus Düllmann Nancy Masschelein
10	2006	The cost efficiency of German banks: a comparison of SFA and DEA	E. Fiorentino A. Karmann, M. Koetter

Visiting researcher at the Deutsche Bundesbank

The Deutsche Bundesbank in Frankfurt is looking for a visiting researcher. Among others under certain conditions visiting researchers have access to a wide range of data in the Bundesbank. They include micro data on firms and banks not available in the public. Visitors should prepare a research project during their stay at the Bundesbank. Candidates must hold a Ph D and be engaged in the field of either macroeconomics and monetary economics, financial markets or international economics. Proposed research projects should be from these fields. The visiting term will be from 3 to 6 months. Salary is commensurate with experience.

Applicants are requested to send a CV, copies of recent papers, letters of reference and a proposal for a research project to:

Deutsche Bundesbank
Personalabteilung
Wilhelm-Epstein-Str. 14

D - 60431 Frankfurt
GERMANY