Macroeconometric forecasting using a cluster of dynamic factor models

Christian Glocker* Serguei Kaniovski†‡

July, 2021

Abstract

We propose a modeling approach based on a set of small-scale factor models linked together in a cluster with linkages derived from Granger causality tests. GDP forecasts are produced using a disaggregated approach across production, expenditure and income accounts. The method combines the advantages of large structural macroeconomic models and small factor models, making our Cluster of Dynamic Factor Models (CDFM) useful for large-scale model-consistent forecasting. The CDFM has a simple structure and its forecasts outperform those of a variety of competing models and professional forecasters. In addition, the CDFM allows forecasters to use their own judgment to produce conditional forecasts.

JEL-Codes: C22, C53, C55, E37

Key Words: Forecasting, Dynamic factor model, Granger causality, Structural modeling

---

*Austrian Institute of Economic Research, Arsenal Object 20, 1030 Vienna, Austria. Email: christian.glocker@wifo.ac.at.

†Austrian Institute of Economic Research, Arsenal Object 20, 1030 Vienna, Austria. Email: serguei.kaniovski@wifo.ac.at.

‡**Acknowledgment**: We would like to thank two anonymous reviewers for thoughtful advice and constructive comments. We also thank Massimiliano Marcellino, Ferdy Adam and his team, and participants at WIFO/Vienna and the 2nd Vienna Workshop on Economic Forecasting 2020 at IHS/Vienna for valuable comments on an earlier draft. We are greatly indebted to Astrid Czaloun for excellent research assistance. All errors are our own responsibility. Financial support by the Austrian Ministry of Finance (Bundesministerium für Finanzen – BMF) is gratefully acknowledged.

1 Introduction

The aim of this paper is to provide a coherent methodology for large-scale macroeconomic nowcasting and forecasting using a rich set of economic indicators available on a monthly basis. To this end, we propose a *Cluster of Dynamic Factor Models* (*CDFM*).

The analysis of an economy in the short term and the projection of its future course are fundamental tasks of central banks, national and international institutions. Two different kinds of models are commonly used in this context: large-scale structural macroeconomic models,[1] and small-scale time series models.[2] Large-scale macroeconomic models typically feature a set of stochastic equations reflecting the behavior of economic agents, supplemented by definitional and institutional relationships (Diebold, 1998). Their heavy reliance on economic theory makes them useful in terms of interpreting forecasts, but at the same time their inability to incorporate information from soft indicators and other high-frequency data renders them less useful for unconditional forecasts. Small-scale time series models attempt to exploit correlations in observed macroeconomic time series, with little recourse to economic theory (Clements and Hendry, 1998). These models are most commonly used to produce unconditional forecasts in a variety of settings, ranging from firm-level business forecasting to economy-wide macroeconomic forecasting.

The popularity of large-scale models stems from the need to forecast many macroeconomic variables and produce conditional forecasts for scenario analysis. While these models allow a model-consistent forecast of many macroeconomic variables to be made jointly, they face several drawbacks. These include the difficulty of incorporating information from soft indicators, the rigid corset imposed on the equations due to the desired consistency with economic theory, and the inability to use time series with mixed frequencies and missing observations (see Brayton et al., 1997; Diebold, 1998, for instance). The last point, in particular, weighs heavily on the usefulness of these models for making unconditional forecasts, especially in the context of short-term forecasts.

The disadvantages of the large-scale models make up the strength of the small-scale models. The small-scale models are characterized by a high degree of flexibility, which

[1]See, for instance, Heilemann and Findeis (2012); Bårdsen et al. (2012); Hammersland and Bolstad Træ (2014); Qin et al. (2007); Wladyslaw (2011); Eitrheim et al. (1999); Brayton et al. (1997).

[2]See, for instance, Stock and Watson (1991); Mariano and Murasawa (2003); Camacho and Pérez-Quirós (2010, 2011); Banbura and Rünstler (2011); Liu et al. (2012); Kuzin et al. (2013); Camacho and García-Serrador (2014).

allows for quick revisions and extensions. However, the biggest advantage – when estimated using the Kalman filter – is the possibility to account for missing observations and mixed frequencies (Mariano and Murasawa, 2003). This makes these models especially useful for nowcasting and short-term forecasting, as shown in Camacho and Pérez-Quirós (2010, 2011). The main drawback of small-scale models is that their limited scope makes it difficult to produce model-consistent forecasts for many macroeconomic variables, thus reducing their usefulness for scenario analysis.

We propose a forecasting framework that combines the advantages of large-scale and small-scale models, thereby mitigating their individual disadvantages. We do this using a cluster of a series of small-scale dynamic factor models. In a first step, we use economic theory and Granger causality tests to identify linkages among the main macroeconomic variables belonging to the production (supply), expenditure (demand) and income accounts of the Systems of National Accounts (SNA). The set of linkages defines the cluster that interfaces the variables. In a second step, we specify small-scale dynamic factor models for each variable in the cluster. Placing all the individual dynamic factor models in the cluster yields a large-scale macroeconomic forecasting model. The composition of a large number of small models allows for a high degree of flexibility, while the linkages enable model-consistent forecasts for many variables. In addition, this approach allows for the determination of conditional forecasts, which makes this setup useful for scenario analysis.

The CDFM features some eighty variables, including variables of the SNA, leading indicators, financial market variables, labor market indicators, price and wage specific variables and a series of variables of key trading partners. These variables are contained in twenty-five individual dynamic factor models. These in turn are linked by the so-called *link-variables* in each individual dynamic factor model.

The forecasting framework we propose allows us to (i) produce model-consistent forecasts for a wide range of variables, (ii) perform scenario analysis by relying on the concept of conditional forecasts, (iii) use mixed-frequency data with missing-observations, (iv) make quick extensions for specific variables and models. Moreover, this approach allows us not only to forecast real-time GDP, but also to include information on the components explaining the forecast, which provides insight into the causes of forecast revisions. Last but not least, our approach allows us to incorporate subjective judgment and thus produce conditional forecasts.

We find that the nowcasting and forecasting performance of the CDFM not only surpasses that of naive models but also that of alternative models that allow for richer dynamics. For example, the CDFM's forecasts for GDP are significantly more accurate than those of classical small-scale dynamic and large-scale factor models applied directly to GDP, as well as several other competing models (e.g., ARIMA). The results are shown to be robust to various model specifications. For example, adding more indicators to our model does not necessarily improve the forecast performance. Last but not least, the forecasts of the CDFM model encompass those of professional forecasters.

Our contribution is related to Banbura and Rünstler (2011), Marcellino and Schumacher (2010), Angelini et al. (2011), Schumacher and Breitung (2008), Barhoumi et al. (2008) and Marcellino and Sivec (2021), among others, who use the approximate dynamic factor model proposed by Giannone et al. (2008) to forecast GDP. Our paper is also related to studies assessing the gains from pursuing a disaggregated (also referred to as *indirect*) approach for forecasting GDP. Pareja et al. (2020) focus on the expenditure account only. They find evidence for a higher predictive accuracy when following an indirect approach. Proietti et al. (2021) propose a method for nowcasting and forecasting sixteen main components of GDP from the production and expenditure account using a large set of monthly indicators and all possible mixed frequency bivariate models of the quarterly GDP components. They find that the indirect approach yields more accurate forecasts for GDP. Heinisch and Scheufele (2018) find evidence to the contrary. They consider the expenditure and production accounts as a means to disaggregate GDP in its components. Esteves (2013) focuses on the expenditure account and finds evidence for a reproducability of the GDP direct forecasts by an indirect approach, provided the same set of indicators is used. Finally, Cobb (2020) considers an indirect approach in forecasting GDP focusing on the production account. He finds that indirect GDP forecasts perform equally well or better than the direct benchmarks. Our contribution differs from these studies along at least two dimensions. First, these studies do not cover all three accounts (production, expenditure and income) of the SNA. Secondly, and most importantly, they ignore the dependency structure among the variables of the SNA that allows to improve the forecasting accuracy.[3] While all these studies rely on a purely statistical approach,

[3]While Cobb (2020) does account for the dependency among the SNA variables, this is confined to

our methodology explicitly takes economic theory into account. The possibly biggest advantage of our approach compared to Pareja et al. (2020), Cobb (2020) and Proietti et al. (2021) is its simplicity and transparency, which makes the proposed method attractive for applied work.

The outline of the paper is as follows. Section 2 introduces the cluster which defines the basic environment of the CDFM. Section 3 describes the individual models, their structure and the basic workings of the overall model. Section 4 provides an extensive model assessment. This concerns both the in-sample fit of each individual model and the CDFM's out-of sample predictive accuracy. Finally, Section 5 concludes.

2 The structure of the cluster

The starting point of our analysis is the SNA. It offers a coherent and consistent set of macroeconomic aggregates for the analysis of the economic structure. These are compiled quarterly according to the methods and definitions outlined in Eurostat (2013).

We use data for Austria, a small open economy, to illustrate the concept of the CDFM. Our set of data comprises the time series of GDP and various components taken from the production, income and expenditure account of quarterly National Accounts compiled by the Austrian National Statistical Agency. The series range from 1996:q1–2019:q4 at chained volumes. The series for the income account are available in nominal terms, i.e. at current prices, only.

Table 1 lists the variables from the three SNA accounts. The summary statistics for all SNA variables can be found in Table 10 of Section F of the appendix. In a first step, we use the concept of Granger causality to identify a structure within this dataset in terms of the variables' mutual predictive contribution.

The Granger-causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another. We run bivariate Granger-causality tests among the year-over-year (y-o-y) growth rates of all variables listed in Table 1. A directional link from variable x_t to variable y_t is established if x_t Granger-causes y_t, but y_t does not Granger-cause x_t. The resulting link is thus unidirectional. It allows us to construct an ordered sequence, or a hierarchy of individual models. Nominal variables

the components of the production account only.

such as, for instance, aggregate labor and capital income from the income account are deflated by using the GDP deflator. We rely on the Bayesian information criterion (BIC) to determine the optimal lag lengths of the individual bivariate models involved in the hypothesis testing. Finally, we used a recently developed multivariate and a nonlinear causality test to check the robustness of the linkages revealed by the standard bivariate Granger-causality test. The results for these alternative tests can be found in Section A of the appendix.

Table 1: National Accounts: Variable coverage

GDP (Production)	GDP (Expenditure)	GDP (Income)
Manufacturing VA (NACE B-E)	Private consumption	Labor income
Construction VA (NACE F)	Investment	Manufacturing
Services VA (NACE G-N)	Construction	Construction
	Equipment	Services
	Intangibles	Capital Income
	Exports	
	Goods	
	Services	
	Imports	
Residual	Residual	Residual

The Cluster of Dynamic Factor Models covers the production, expenditure and income side of the quarterly National Accounts, as well as employment and other monthly indicators.

In an attempt to identify a plausible structural cluster among the variables in Table 1, we supplement the Granger-causality tests with economic theory.[4] For this we proceed in two steps. The first involves the identification of Granger-causal dependencies. Here we assign a level of statistical significance of ten percent within the Granger-causality tests. If a dependency between two variables, as identified by means of the Granger-causality test, is at odds with economic theory, we discard this linkage. While this intervention was not necessary in our case for the dependencies identified at the five percent level or

[4]The structure of the production account depicted in Table 1 has a very basic form. However, it could easily be extended to take into account country specific production characteristics. In case of a typical raw-material goods producing country, one could decompose the value added in manufacturing into further sub-categories; the same applies to the value added in the service sector. The flexible structure of the CDFM allows for extensions along various dimensions in this context.

higher, it becomes important at the ten percent level.[5] The results of this exercise are depicted as arrows in Figure 1. We show the results in the form of a graph. The arrows replicate the bivariate Granger-causality test results and hence identify the linkages that are important for improving a forecast from a statistical point of view.

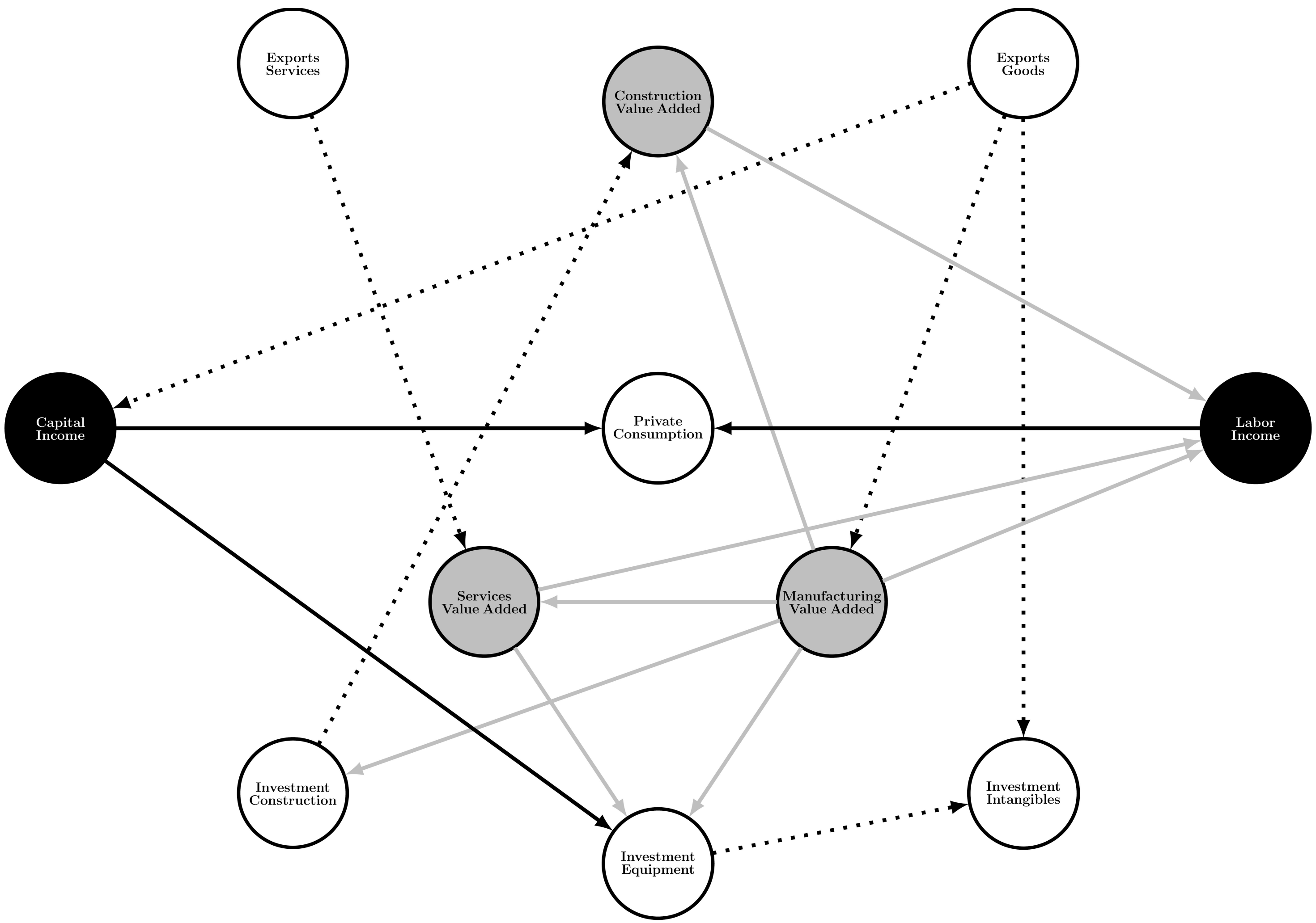
The results conform with the intuition for a typical small open economy. Exports (of goods and services) are the key variable in driving the dynamics. Shocks therein immediately affect the manufacturing and service sectors, which in turn provides incentives for entrepreneurs to expand on investment (equipment and construction investment). The increasing importance of digital elements in equipment investment motivates equipment investment as a means of stimulating intangible investment (intangible investment primarily involves spending on computer software, etc.). These linkages identify dependencies between the variables of the production and expenditure accounts and shape the dynamics of these variables and hence those of the economy as a whole.

On the income account, we observe that labor income is solely affected by variables of the production account (value added in manufacturing, construction and services). Capital income is in turn only affected by exports, highlighting once more the essential role of exports in shaping the overall income path. Changes in both labor and capital income transmit to private consumption. While labor income only affects consumption, capital income also affects equipment investment. The linkage arises because, from an entrepreneur's perspective, capital income expands the equity base. This reduces the level of indebtedness, which facilitates and favors borrowing, which ultimately stimulates investment. This link is statistically different from zero at the one percent level, highlighting the importance of capital income in shaping investment dynamics.

While we interpreted the arrows shown in Figure 1 as causal above, this is done only to explain them in an economic context. Granger causality tests do not necessarily allow for a causal interpretation, as they only evaluate the informational content of the variables for the purpose of forecasting (Hamilton, 1994).

[5] We checked our final results for these omitted linkages, and we find that neither of these omissions has the potential to improve the forecast accuracy of the CDFM.

Figure 1: Granger-causal links



Note: Variables in white circles are from the expenditure account, the ones in black circles from the income account and the ones in gray circles from the production account.

The structure of the production account proposed here is commonly considered by macroeconomic forecasters. While it comprises a reasonable approach from the point of view of a practitioner, this structure nevertheless relies only on the classification of the production account within the SNA, but it is not an approach proposed by economic theory. Against this background, we also assess the CDFM when using a different approach in modeling the production account. We do so by relying on a distinction between tradable and nontradable goods and services (TNT). Further details are provided in Section B of the appendix.

In what follows, we incorporate the Granger-causal linkages within a series of small-scale dynamic factor models. We specify individual models for each variable depicted in Figure 1 and Table 1. The linkages constitute the structural element of the cluster, with the small-scale dynamic factor models being a reduced form element. The result is a *structural* cluster of dynamic factor models. While the linkages depicted in Figure 1 allow only for a limited degree of dependency between the variables, this nevertheless proves sufficient for establishing model-consistent forecasts for a wide range of variables.

3 The cluster of dynamic factor models (CDFM)

The basic element of the CDFM is the dynamic factor model (DFM). A DFM provides a parsimonious representation of macroeconomic data, where few factors explain the majority of co-movements among macroeconomic data series (see Camacho and Pérez-Quirós, 2010; Foroni and Marcellino, 2014; Banbura and Rünstler, 2011; Rusnák, 2016; Jiang et al., 2017; Liu et al., 2012, among others). The cluster of dynamic factor models (CDFM) comprises a series of dynamic factor models (DFMs), each estimated using the Kalman filter. Link variables across the DFMs connect the individual models to each other.

3.1 Some preliminaries

Let $\boldsymbol{x}_t = [x_{1,t}, \ldots, x_{n,t}]^\intercal$, $t = 1, \ldots, T$, denote a set of standardized stationary monthly variables. Specifically, $\boldsymbol{x}_t$ will be a collection of monthly data taken either in levels or monthly year-on-year (y-o-y) growth rates. To incorporate quarterly y-o-y data (mostly from the National Accounts, e.g. GDP, investment, etc.), we construct a partially observable monthly y-o-y series and link it to the monthly variables by applying a modification of the approximation in Mariano and Murasawa (2003). In what follows, we adopt the convention that time indices for the quarterly variables refer to the third month of each quarter. Following Modugno et al. (2016); Kuck and Schweikert (2021), we consider quarterly level data for a given quarter to be the sum of monthly unobserved contributions. In particular, let X_t^q be a quarterly variable (in log-level), X_t^m its monthly (unobserved) counterpart, and let $x_t^{m,y}$ denote its unobserved monthly y-o-y growth rate. The monthly unobserved y-o-y growth rate can then be linked to a partially observed (at every third month of the quarter) quarterly y-o-y growth rate $x_t^{q,y}$, as follows:

$$
\begin{aligned}
x_t^{q,y} &= X_t^q - X_{t-12}^q \\
&= (1 - L^{12})X_t^q \\
&\approx (1 - L^{12})(1 + L + L^2)X_t^m \\
&= (1 + L + L^2)x_t^m \\
&= x_t^{m,y} + x_{t-1}^{m,y} + x_{t-2}^{m,y}
\end{aligned} \tag{1}
$$

where L is the lag operator. Viewing equation (1) as a factor model implies that quarterly variables should load equally on the current and lagged values of the unobserved monthly growth rate. We apply this set-up for all DFMs. More important, however, is the decomposition of the vector $\boldsymbol{x}_t$ in terms of its role for the cluster. The cluster contains a series of DFMs of which each single DFM (i) addresses a specific target variable and (ii) establishes linkages with other DFMs.

The specification of a DFM for some target variable $x_t^{(j)}$ features an n dimensional vector $\boldsymbol{x}_t^{(j)}$ of observed monthly or quarterly time series which is partitioned as follows:

$$
\boldsymbol{x}_t^{(j)} = \begin{bmatrix} x_t^{(j)} \\ \boldsymbol{x}_t^l \\ \boldsymbol{x}_t \end{bmatrix}
\begin{array}{l} \text{– target variable} \\ \text{– link variables} \\ \text{– other variables} \end{array} \tag{2}
$$

where the vector $\boldsymbol{x}_t^l$ defines the link variables and $\boldsymbol{x}_t$ is a vector of other variables useful for forecasting. To avoid circular dependency structures, we define the vector $\left\{\boldsymbol{x}_t^{*,(j)} = \left[x_t^{(j)}, (\tilde{\boldsymbol{x}}_t)^\intercal\right]^\intercal \mid t = 1, \ldots, T; \tilde{\boldsymbol{x}}_t \subset \boldsymbol{x}_t\right\}$ of dimension p, where $p \leq n$. It is important to note that $\boldsymbol{x}_t^{*,(j)}$ and $\boldsymbol{x}_t^l$ do not have any elements in common. The link variables $\boldsymbol{x}_t^l$ connect the DFM for the target variable $x_t^{(j)}$ to previous DFMs. Inversely, this DFM also passes a subset of variables of $\boldsymbol{x}_t^{*,(j)}$ to subsequent DFMs. In principle, our approach could take circular dependency structures (simultaneity) into account, but this would increase computational complexity significantly. Hence, the link variables of any behavioral DFM only involve variables of previous DFMs, though not of subsequent ones. This implies that the order of the DFMs matters in the CDFM.

Our model assumes that the indicators are generated according to a stationary model, such that $\boldsymbol{x}_t^{(j)}$ has an approximate factor structure, in the sense specified by Forni et al. (2000) and Forni and Lippi (2001). Our disaggregated modeling approach requires the aggregation of some variables at several points in order to account for identities of the SNA. To this purpose, we distinguish between *behavioral* models and *aggregator* models. Next, we discuss each of them in detail.

3.2 A behavioral DFM

We identify *behavioral* models to explain the dynamics of particular variables of interest. Consider again $x_t^{(j)}$ as a target variable. We specify a small-scale dynamic factor model using a q-dimensional, $0 < q < n$, vector of factors $\mathbf{f}_t$:

$$\boldsymbol{x}_t^{(j)} = \boldsymbol{\Lambda}(L)\mathbf{f}_t + \boldsymbol{D}(L)\boldsymbol{\epsilon}_t \tag{3}$$

$$(\boldsymbol{I} - \boldsymbol{\Phi}(L))\mathbf{f}_t = \boldsymbol{e}_t \tag{4}$$

where $\boldsymbol{\Lambda}(L)$ are $n \times q$ loading matrices which take into account equation (1). The common component $\boldsymbol{\Lambda}(L)\mathbf{f}_t$ and the idiosyncratic term $\boldsymbol{D}(L)\boldsymbol{\epsilon}_t$ are assumed to be uncorrelated, and, moreover, $\boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{\Sigma})$, such that $\boldsymbol{\Sigma}$ is an $n \times n$ diagonal covariance matrix. The matrix $\boldsymbol{\Phi}(L)$ is a lag-polynomial governing the dynamics of the latent factors in $\mathbf{f}_t$. The error term $\boldsymbol{e}_t$ of the dynamic equation satisfies: $\boldsymbol{e}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_e)$. For identification reasons, we impose that $\boldsymbol{\Sigma}_e$ is equal to the identity matrix.

We cast equations (3) and (4) into a state space format and estimate its parameters by running the Kalman filter. We set up the Kalman filter to deal with missing observations, as discussed in Durbin and Koopman (2001). A sequence of such models is estimated individually. We standardize each element in the vector $\boldsymbol{x}_t^{(j)}$. This is advisable because it reduces the number of parameters to be estimated and homogenizes variances of the idiosyncratic components which, in turn, allows for a significant acceleration of the estimation process of the models.

The specification of the behavioral models involves a decision on (i) which variables to include and in case a particular variable is included, (ii) the temporal displacement (contemporaneous, lagged, etc.). We do so by relying on a combinatorial algorithm (Glocker and Wegmüller, 2020), that targets the out-of-sample forecasting accuracy (average of the root-mean-squared error (RMSE) of the first three quarters) of the target variable $x_t^{(j)}$ as the objective[6]. Since our sample is small, we do not consider a presample to perform variable selection. Moreover, the selection does not concern the link variables, since these are come from the results of the Granger causality test.

In the course of this selection process,[7] we find that for most target variables comparatively small models already have a forecast precision that cannot be further improved when including additional variables.[8] This results in a multitude of small-scale models. The big advantage here is that the estimation process of small models is fast in the context of a non-linear optimization routine, as is the case with the Kalman filter, and the variables can be adapted optimally with respect to their temporal displacement relative to the target variable.

The first column in Table 11 gives an overview of all the variables for which we specify a behavioral DFM. They are referred to as target variables and captured by $x_t^{(j)}$ in the vector in equations (2) and (3). The variables involve those mentioned in Figure 1, but for several variables we consider a more disaggregated approach. This applies, for instance, to labor income, where we distinguish between labor income arising from the manufacturing, construction and service sector. The ordering of the variables follows a specific pattern:

[6]In order to accelerate the process of variable selection, we incorporate the findings from Giannone et al. (2008), Heinisch and Scheufele (2018), Lehmann (2021), Klein and Özmucur (2010) among others.

[7]The above procedure enables several options. For example, the RMSE considered is based on the average of the RMSEs of the first three quarters. A horizon different from this would naturally lead to a different set of selected variables and a different temporal displacement. Moreover, the set of variables from which the algorithm selects includes only *timely available* (monthly) indicators. Adding indicators with a larger release lag would also result in a different set of selected variables. With this in mind, we evaluated the sensitivity of each behavioral model with respect to the variables selected by the algorithm. It was found that while the behavioral models' forecast accuracy (RMSE) changes slightly, the results of the Diebold-Mariano test remained unchanged.

[8]As described in Boivin and Ng (2006) and Banbura and Rünstler (2011), the inclusion of additional variables, despite possible high correlation with the target variable, does not necessarily improve the forecast. When an additional variable is correlated with a subset of variables already in the model, the factors have a bias towards this subset of variables. As a consequence, the resulting factors explain a large fraction of the variation in each variable of this subgroup, but less of the variance in the target variable, rendering worse the overall model fit for the target variable and hence also its forecast.

from the most exogenous (foreign variables) to progressively more endogenous variables (pure domestic variables). This sequence replicates the empirically observed dependence structure and thus constitutes a key feature of the CDFM. The second column provides a list for the link variables ($\boldsymbol{x}_t^l$) for each behavioral DFM. The link variables replicate the arrows depicted in Figure 1 and hence connect the individual DFMs to each other. Finally, the third column in Table 11 lists a series of additional variables. These are used to improve upon the forecast of the target variable of the individual DFMs. Their selection as well as their temporal displacement is based on their contribution to improving the out-of-sample forecast, for which we rely upon the root-mean-squared-error (RMSE).

3.3 An aggregator DFM

While the behavioral DFMs identify appropriate reduced form models for the target variables for the purpose of forecasting, they, however, exclude aggregate variables. This arises from our modeling approach, which is highly disaggregated. In order to forecast aggregate variables, we therefore construct *aggregator models*. Aggregator models essentially rely on identities. Consider the computation of GDP along the production account. The sum of the value added of all sectors yields the total value added and, once taking into account taxes and subsidies on products, we end up at GDP by definition. This is an identity for which we consider only a selected number of subcomponents within our approach. From this, two problems arise: first, modeling levels or growth rates, and second, how to deal with omitted components? We proceed in two steps. In the first, we consider a log-linearization of this identity. This gives us the GDP growth rate as a weighted average of the growth rates of each subcomponent. The weights are given by each subcomponent's share in GDP. This allows us to continue working with growth rates instead of levels. In the second step, we add a residual term. This term captures (i) changes in the shares of the subcomponents in GDP over time and (ii) collects all those subcomponents which are not specifically addressed by behavioral DFMs.[9]

More formally, an aggregator model for y_t involves a weighted sum of the constituent component series. Let the series y_t comprise r components $x_t^{(i)}$ for $i = 1, \ldots, r$. An

[9] The error term serves one further, more subtle, aspect: as highlighted in Proietti et al. (2021). The weighted sum of the yearly growth rates does not add up to GDP growth and can only be seen as an approximation to the contribution of the ith component to aggregate growth.

aggregator model for y_t reads:

$$y_t = \sum_{i=1}^{r} \omega_i x_t^{(i)} + \theta(L)\eta_t \tag{5}$$

$$(1 - \varphi(L))(\eta_t - \mu) = \epsilon_t \tag{6}$$

where $x_t^{(i)}$ $\forall i = 1, \ldots, r$ and y_t are non-standardized growth rates, $\theta(L)$ and $\varphi(L)$ are lag-polynomials, μ is a constant term and the error term ϵ_t satisfies $\epsilon_t \sim N(0, \sigma_\epsilon^2)$. The weights ω_i $\forall i = 1, \ldots, r$ are fixed at their most recently observed values. We proceed as in Section 3.2 and cast equations (5) and (6) into a state space format and estimate its parameters using the Kalman filter.

We proceed in this fashion to establish aggregate growth rates for GDP along the production, expenditure and income accounts. Further aggregator models are established for labor income, employment, investment, exports and imports. Table 12 provides an overview for each aggregator model and its subcomponents.

3.4 How the CDFM works

The key element for the working of the model is the concept of conditional forecasts. Conditional forecasting concerns forecasts of endogenous variables conditional on a pre-determined path of endogenous variables. Specifically, it is assumed in our context that the conditioning information satisfies *hard conditions* (a particular path) rather than *soft conditions* (a range for the path). This stands in contrast to unconditional forecasts, where no knowledge of the future path of any variable is assumed.

The conditional forecasts of any DFM are passed on to another DFM within the cluster. To see how this works in practice, consider the following. For a small open economy, the most relevant shocks usually emerge from foreign demand. Therefore let us start with the DFM for goods exports (DFM (3)). This model can be used to produce either conditional or unconditional forecasts. In this context, a reasonable conditional forecast for goods exports could arise from assuming a certain future GDP path of the foreign economies. In case no future path of the foreign economies is assumed, the DFM for goods exports will establish an unconditional forecast for all the variables in the vector $\boldsymbol{x}_t^{(j)}$ of the DFM for goods exports. From this vector, the unconditional forecast of several variables is in turn used in yet other DFMs. The passing on is established by means of the link variables ($\boldsymbol{x}_t^l$). This concerns, among others, the use of foreign PMIs for the model for service exports (DFM (4)). These linkages ensure consistency of the external environment in the models for the exports of goods and services. Furthermore, the unconditional forecast of goods exports enters the DFM for the value added in the manufacturing sector (DFM (5)), capital income (DFM (12)) and intangible investment (DFM (16)). Once the forecast of goods exports is used in these models, conditional forecasts are established for the value added in the manufacturing sector and capital income. These conditional forecasts are in turn used in further downstream DFMs.

This shows the role of conditional forecasts as a means of operationalizing the linkages within the CDFM. While the previous example started from an unconditional forecast for goods exports, the exercise would essentially be unchanged when a given future path of GDP for the foreign economies is passed as a conditional forecast of goods exports to the subsequent DFMs.

In principle, the joint reliance on unconditional and conditional forecasts renders an assessment of the adequacy of our model difficult. For instance, one can immediately see why the failure of unconditional forecasts tells us little about how good a model is at conditional forecasting. A macroeconomic model may be reasonably good at predicting, for instance, how a change in oil prices will influence output, but it can still be poor at predicting what output growth will be next year, because it is inadequate at predicting oil prices in the first place. In this context, the CDFM offers the possibility of substituting imprecise unconditional forecasts of the CDFM for a specific variable with forecasts from outside. The CDFM then establishes forecasts for all variables in the model, conditional on the specific path assumed for this variable. The CDFM hence offers a flexible environment, where inadequate forecasts can quickly be adjusted. This in turn also allows for scenario analysis in order to assess the sensitivity of forecasts to changes in the variables of some preceding DFMs. We provide further details in Section 4.7.

3.5 Consistency

The CDFM computes forecasts for GDP along three dimensions: production, expenditure and income account. This seemingly independent three-fold approach might cast doubts on the predicted values once they diverge. We therefore assess the extent to which the GDP predictions arising from the three accounts are consistent.

The CDFM links variables from the three accounts to each other. Consider, for instance, investment as one particular expenditure variable. Its subcomponents (equipment investment, construction investment and intangible investment) are Granger-caused by variables from both the production account (manufacturing value added) and the income account (capital income). Variables from the expenditure account in turn also Granger-cause variables of the production and income accounts. A multitude of inherent linkages of this kind is likely to contribute to the overall model consistency, as shocks in a variable are transmitted across all three accounts via these linkages. Since discrepancies between the aggregates can still arise, it is important to compare the discrepancies from the CDFM with the empirically observed counterparts.

We consider two empirically observed counterparts: (i) data revisions and (ii) inventory investment (we consider inventory investment[10] jointly with the official statistical discrepancy). Concerning the first, the data of the SNA are continuously revised. Revisions usually result in large changes of past growth rates. As regards the second, while for most countries the primary approach in computing the GDP is based on the production account, the expenditure and income accounts are adjusted accordingly. In principle, a GDP figure can be established from each of the three accounts. Since discrepancies are likely to arise, statistical authorities in general use the subcomponents to correct for them. These subcomponents are inventory investment for the expenditure account and capital income for the income account. In this respect, the CDFM closely mimics the approach and the difficulties statistical agencies face when computing the GDP within the SNA.

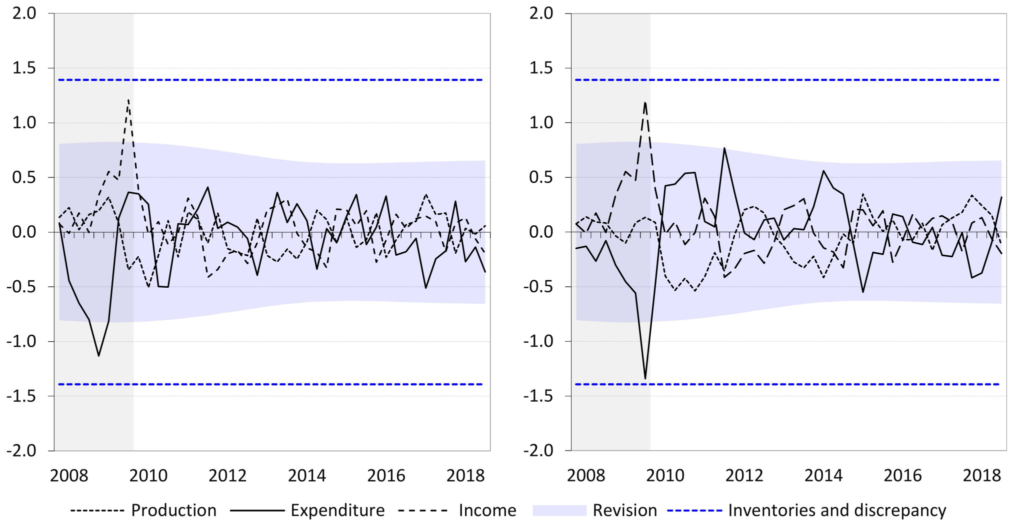
We compare the discrepancies of the CDFM for GDP from the three accounts (production, expenditure and income) to empirically observed revisions in the official data and differences in the GDP growth figures arising from inconsistencies across the three accounts. Figure 2 shows a band for the average revisions[11] and a measure for the empirically observed discrepancies in the growth rates across the production and expenditure account[12] depicted by the blue dashed line. In addition to this we show the one-quarter

[10]This refers specifically to changes in inventory investment and acquisitions less disposals of valuables.

[11]Average revisions are computed as the standard deviation across each quarter; this is possible since for each quarter, various estimates are available (flash estimate, first, second, ..., official figures; in our case for some quarters we have up to fifteen distinct figures for the quarterly growth rates). The blue band shows the (smoothed) average revisions and the change thereof over time. The extent of the data revisions depends, among other things, on the level of the growth rates. Although revisions are on average zero, their standard deviation is large compared to the mean of the GDP growth rate.

[12]We compute the GDP growth contribution of inventory investment and the statistical discrepancy

Figure 2: Internal consistency



Note: Internal consistency at 3m(1q) (left) and 12m(4q) (right). The values are in percentage points.

and four-quarter ahead forecast errors of the CDFM for GDP growth from all three accounts. The GDP growth rates from the CDFM from the production, expenditure and income account differ, but the discrepancy is small compared to the average of the empirically observed discrepancy (around 1.5 percentage points over the horizon 2008–2018). This applies to both the one-quarter and four-quarters ahead forecasts. More importantly, however, is the observation that the discrepancies implied by the CDFM are noticeably smaller than those arising from the revision of official figures. The extent of official data revisions surpasses the discrepancies implied by the CDFM to a large extent, with only two exceptions in case of the one-quarter ahead forecast and three exceptions within the four-quarter horizon of the CDFM forecast.

When looking at the second measure of comparison – the standard deviation of the difference in actual growth rates – we find an even larger divergence in favor of the CDFM. The average gap between GDP growth as of the production and expenditure accounts from the official figures surpasses the discrepancies of the CDFM by a factor of up to four. We conclude that the GDP forecasts from the CDFM arising from the production, income and expenditure accounts differ to some extent. This difference, however, is small when

using the final data vintage. We then calculate the standard deviation of the growth contributions. The value thereof is depicted by the blue dashed line in Figure 2. The standard deviation is shown in positive and negative territory in order to establish an interval.

compared to official figures from the SNA. We interpret this result in favor of a model inherent consistency.

In principle, our approach could easily be extended to obtain a single GDP measure across the three accounts, for example by looking at the average across the three accounts, which we will do later. Another possibility is given by the error terms (η_t in equation (5)). Since all aggregator models for GDP include an error term, discrepancies could always be cast into these error terms to ensure that only a single overall GDP measure emerges from the CDFM. This, however, requires the application of further statistical procedures as, for instance, described in van der Ploeg (1982). In this context, Pareja et al. (2020) impose a balancing procedure which allows consistent forecasting of macroeconomic aggregates through an equilibrium model. Their approach differs from ours in that they model GDP and its components separately. Discrepancies contained therein occur due to a lack of consistency between individual DFMs, as they are not connected to each other, which requires a balancing procedure.

4 Results

We start an assessment of the CDFM with an in-sample analysis. To this purpose, we consider the behavioral DFMs only and use the R^2 as a measure of fit in this context. It is computed by regressing the factors of the jth behavioral DFM on the target variable $x_t^{(j)}$. The results can be found in the second column in Table 2.

The in-sample fit ranges from a low of 0.52 in the case of the models for the value added of services and construction up to 0.98 in case of intangible investment. Across all models the average value of the R^2 is around 0.80. This value conforms with those reported in other studies. Glocker and Wegmüller (2020), for instance, identify a small-scale DFM for Switzerland; their preferred model has an R^2 of 0.74. In the specification of Camacho and Pérez-Quirós (2011) the factor explains a share of 0.8 of the variance of GDP growth. While these numbers seem sufficiently high for the models to produce good forecasts, this does not necessarily have to be the case. One reason for this is that a model with a good in-sample fit does not necessarily produce a precise out-of-sample forecast (Clark, 2004; Granger and Jeon, 2004). The opposite also holds owing to the dynamic structure of the error term ($\boldsymbol{D}(L)\boldsymbol{\epsilon}_t$) that improves the forecast beyond the contribution of the factors $\boldsymbol{f}_t$.

All this can be seen when considering the out-of-sample forecasts to which we turn next.

Table 2: Factor correlation and NRMSE by behavioral DFM (2007-2018)

		in-sample (R^2)	out-of-sample ($NRMSE$)			
DFM	Variable		3m(1q)	6m(2q)	9m(3q)	12m(4q)
(1)	Import deflator	0.77	0.44	0.86	1.14	1.29
(2)	Private consumption deflator	0.60	0.38	0.61	0.81	0.97
(3)	Export of goods	0.78	0.41	0.74	1.04	1.26
(4)	Export of services	0.57	0.57	0.59	0.56	0.63
(5)	Manufacturing VA	0.93	0.54	0.75	1.01	1.15
(6)	Investment construction	0.52	0.46	0.63	0.76	0.85
(7)	Construction VA	0.73	0.53	0.79	1.04	1.20
(8)	Services VA	0.52	0.44	0.64	0.79	0.94
(9)	Labor income manufacturing	0.83	0.38	0.59	0.82	1.04
(10)	Export deflator	0.97	0.43	0.85	1.15	1.31
(11)	Capital income	0.92	0.65	0.87	1.05	1.16
(12)	Labor income construction	0.75	0.55	0.62	0.69	0.75
(13)	Labor income services	0.87	0.38	0.58	0.79	0.96
(14)	Private consumption	0.56	0.75	0.88	0.95	0.98
(15)	Investment equipment	0.73	0.66	0.76	0.86	0.91
(16)	Investment intangibles	0.98	0.51	1.12	1.79	2.27

NRMSE refers to the normalized (standard deviation) root-mean-squared-error and hence ensures the comparability across all variables. The largest forecast errors are observed for consumption, capital income and equipment investment.

4.1 Out-of-sample analysis

We construct a real-time dataset to assess the forecasting performance. Since we rely on final data rather than data vintages, we consider this approach to be a pseudo real-time analysis.[13]

We construct our pseudo real-time dataset on monthly vintages. For each month within the 2006-2019 period we collect the whole set of time series available. We end up with 156 different vintages for the period 01/2006 to 12/2019. Our pseudo-real time

[13]In contrast to Camacho and Pérez-Quirós (2011) and others, we ignore data revisions of the variables of the SNA. This arises primarily because we do not have historical data records for the majority of the variables in the SNA; and for the remaining variables, the time span of historical data records is too short to plausibly conduct an evaluation of forecast accuracy taking into account data revisions.

analysis replicates the idea suggested in Proietti et al. (2021), that is we use the ragged structure of monthly indicators at the end of each month to establish a forecast. This dataset allows us to closely mimic the forecasting procedure a practitioner would have performed at any point in time when computing forecasts.

4.2 Predictive accuracy: a general view

We rely on the root-mean-squared-error (RMSE) as a measure of accuracy of quarterly out-of-sample forecasts. To ensure comparability of error measures over a wide range of variables, we normalize the RMSE by the standard deviation of the same variable. The normalized measure is referred to as a normalized-root-mean-squared-error (NRMSE). We provide its values for the target variables specified in behavioral DFMs in the third to fifth columns of Table 2. Table 13 of the appendix provides the results for the mean-absolute-error (MAE). Across all variables, we find that the NRMSE increases with the horizon, which implies that the out-of-sample forecasting precision declines with the horizon. We find that the DFM for the consumption deflator (DFM (2)) and labor income in the manufacturing sector (DFM (9)) and in the service sector (DFM (13)) yield the most precise one-quarter-ahead forecasts. For the four-quarter ahead horizon, the DFM for service exports (DFM (4)) has the highest forecast accuracy. While this provides some information on the relative adequacy of each individual DFM, the comparison with the in-sample fit is, however, more interesting. In this context, the DFMs for service exports and intangible investment stand out. While the DFM for service exports has a rather low R^2, it at the same time has the highest forecast precision at the four-quarter horizon. We observe the opposite in the case of the DFM for intangible investment: with an R^2 of 0.98, this model has the highest in-sample fit; however, its four-quarter ahead forecast accuracy is the worst across all models. This once more highlights the commonly found observation that a model with a good in-sample fit does not necessarily produce accurate out-of-sample forecasts.

We provide the values of the NRSME for the variables which are captured by means of aggregator models in Table 3 and the results for the MAE are provided in Table 14 of the appendix. As can be seen exports and labor income have a comparably high forecast accuracy at the one-quarter horizon. The high precision of the model for labor income stands out at the four-quarter horizon.

Table 3 also provides NRMSE values for the GDP measures. Since the CDFM allows to establish GDP figures along the production, expenditure and income account jointly, we can hence assess the forecast accuracy for GDP from each of these three accounts. We add a fourth measure for GDP, which is simply the (unweighted) average across the three former measures. We find that the GDP forecast accuracy is highest with respect to the expenditure account, followed by the production account. The GDP forecast of the production account could be noticeably improved when relying on the concept of tradable and nontradable goods and services (see Section B of the appendix). We find that the GDP forecast from the income account has the lowest precision. This, however, only applies to the one-quarter horizon. Considering the four-quarter horizon, we find that the highest forecast precision for GDP now emerges from the income account. The average GDP measure performs reasonably well: worse than the expenditure approach, but better than the other two at a short horizon. For the four-quarter horizon, the average GDP forecast now outperforms the expenditure account.[14]

We compare the forecast accuracy of the CDFM with a series of competing models. These are: (i) a random-walk model, (ii) an AR(1) model, (ii) an optimal ARIMA model, (iv) a small-scale dynamic factor model directly applied to GDP, (v) a large-scale dynamic factor model involving all variables that are part of the CDFM and finally (vi) a MIDAS regression model as suggested in Kuzin et al. (2013). The optimal ARIMA model is specified by relying on the Bayesian information criterion (BIC) for the lag-lengths of the autoregressive and moving-average lag-polynomials. Details on the specification of the small-scale and large-scale dynamic factor models and the MIDAS regression model can be found in Sections C – E of the appendix.

The out-of-sample forecasting performance of the competing models is also shown in in Table 3, where we again use the normalized RMSE (NRSME). As can be seen, the competing models have noticeably higher values for the NRMSE for both short and long forecasting horizons compared to the CDFM. The small DFM and the MIDAS regression models applied directly to GDP notably outperform the other competing models. Moreover, for a short horizon they perform as well as the GDP forecast of the CDFM arising

[14]The average GDP forecast is forecast combination. In this context Moser et al. (2007); Kapetanios et al. (2008); Öğünç et al. (2013); Foroni and Marcellino (2014) argue that, in general, a forecast combination reduces forecast errors.

from the income account.[15] Finally, the CDFM tends to outperform the large-scale DFM, especially for short horizons.

Table 3: NRMSE by aggregator DFM (2007-2018)

DFM	Variable	3m(1q)	6m(2q)	9m(3q)	12m(4q)
(17)	Exports	0.36	0.65	0.91	1.14
(18)	Imports	0.52	0.79	1.04	1.26
(19)	Investment	0.85	0.90	1.02	1.15
(20)	Labor income	0.36	0.55	0.76	0.95
(21)	Employment	0.39	0.67	0.90	1.09
(22)	GDP deflator	0.57	0.73	0.76	0.76
(23)	GDP production	0.43	0.62	0.86	1.04
(24)	GDP expenditure	0.39	0.62	0.89	1.09
(25)	GDP income	0.50	0.70	0.90	1.03
	GDP average	0.41	0.61	0.86	1.05
			Competing models		
	GDP random walk	0.60	0.96	1.28	1.58
	GDP AR(1)	0.58	0.87	1.09	1.27
	GDP ARMA(2,1)	0.56	0.82	1.02	1.18
	GDP Small DFM	0.50	0.76	0.96	1.14
	GDP Large DFM	0.60	0.73	0.85	0.93
	GDP MIDAS	0.49	0.79	0.99	1.12

NRMSE refers to the normalized (standard deviation) root-mean-squared-error and hence ensures the comparability across all variables. The GDP average denotes the mean forecast of the three GDP National Account concepts.

4.3 Predictive accuracy: the importance of the linkages

Since the linkages between the individual DFMs are a fundamental feature of the CDFM, we will now briefly demonstrate their relevance in terms of the gain in the out-of-sample predictive accuracy for GDP.

Tables 2 and 3 show the values of the NRMSE for different horizons based on the

[15] Motivated by the findings in Dias et al. (2015), we analyzed the forecast accuracy of the CDFM for the periods before and after the Great Recession. We find only small differences in the quantitative accuracy measures between the two periods. However, due to the short length of the respective time series for the forecast evaluation, these results should be interpreted with caution. Still, we also interpret this result in favor of our assumption for the time invariant variances in equations (4) and (6).

CDFM involving the linkages shown in Figure 1. We repeat this forecasting exercise based on a version of the CDFM, in which all these linkages have been removed. This means that each individual DFM now produces forecasts independently to those of other DFMs. Therefore, for all those variables that are included in several DFMs, multiple forecasts are produced, which are all different to each other.

Table 4: Forecast error inflation without linkages (2007-2018)

Variable	3m(1q)	6m(2q)	9m(3q)	12m(4q)
Exports	1.00	1.00	1.00	1.00
Imports	1.46	1.18	1.04	0.93
Investment	1.18	1.24	1.18	1.09
Consumption	1.05	1.07	1.04	1.05
Capital income	1.06	1.10	1.02	1.02
Labor income	1.00	0.98	0.95	0.91
Employment	1.21	1.22	1.20	1.14
GDP deflator	1.02	1.01	1.05	1.03
GDP production	1.09	1.11	1.03	1.01
GDP expenditure	1.51	1.26	1.10	1.05
GDP income	1.08	1.14	1.11	1.07
GDP average	1.15	1.16	1.08	1.04

We compare the predictive accuracy of this alternative model with the CDFM. Table 4 relates the values of the NRMSE of the alternative model (no linkages) to those of Table 3. Values greater than unity imply that the CDFM's predictions are superior to those of the alternative model (no linkages). The values of the relative NRMSE are always larger than unity. The differences are considerable, implying an improvement of up to 51 percent in the forecast accuracy. We conclude that the CDFM's forecasts are superior to the alternative model without the linkages. This underscores the importance of the linkages for the predictive accuracy of the CDFM. The second purpose of including the linkages is to ensure the internal consistency of a model solution, which is particularly important for forecasting, but also for simulations.

4.4 Predictive accuracy: a closer look

While the results in Table 3 already hint at the predictive accuracy of different models, the question concerning whether some models are systematically better than others is, however, yet left unanswered. To this purpose, we use a modified version of the Diebold-Mariano test (Diebold and Mariano, 1995) as proposed by Harvey et al. (1997). The results of the modified Mariano-Diebold test based on the NRMSE are provided in Table 5 and those based on the MAE in Table 15 of the appendix.

With a view to the CDFM only, several results emerge: (i) the accuracy of the GDP forecasts of the CDFM along the production, expenditure and income accounts is statistically equivalent. For any horizon there is no evidence of a difference at any level of statistical significance. This once more highlights the consistency of the CDFM, as already highlighted in Section 3.5, though with a different approach. (ii) The tradable-nontradable goods approach for modeling the production account as motivated in Section B of the appendix also produces forecasts as good as the basic version of the CDFM. Again, there is no evidence of a difference for any level of statistical significance for any forecasting horizon. While this result shows that different approaches to modeling the production account yield similar levels of forecast accuracy, it once more underscores the consistency inherent to the CDFM despite having three (four) measures for GDP.

When comparing the CDFM with the competing models, several further interesting results emerge: (iii) the CDFM forecasts are more accurate than those of the random walk and the AR(1) model for any horizon. This applies to the CDFM model in general, and to its prediction for GDP along the production and expenditure accounts in particular. (iv) The predictions of the ARIMA model tend to be worse than those of the CDFM for any horizon, and the difference thereof is significantly different from zero. This observation underscores the usefulness of the information contained in the monthly indicators, rendering the CDFM's forecast more precise at any horizon. (v) The CDFM's forecast from the production and the expenditure accounts tends to be superior to the forecast of the small-scale DFM applied to GDP directly. The difference in the precision of the forecast is statistically significant at the one percent level for the one-quarter and two-quarter horizons. For larger horizons, the GDP forecasts of the CDFM are on average more precise, but the difference is not statistically different from zero. (vi) The small-scale DFM's forecasts are more accurate than those of the random-walk (all horizons) and the AR(1) and ARIMA models (one-quarter and two-quarter ahead horizon). This result replicates the findings in Camacho and Pérez-Quirós (2010), Camacho and Pérez-Quirós (2011) and Camacho and García-Serrador (2014), to mention a few. (vii) The CDFM's forecast are superior to those of a large-scale DFM for short horizons (one quarter and two quarters ahead); however, for longer horizons (four quarters and beyond), the forecasts of the large-scale DFM turn out to be more accurate. Without examining this in more detail, a possible reason for this might be the quarterly frequency considered in the large-scale DFM, whereas the CDFM is specified at a monthly frequency. Finally, (viii) the MIDAS regression model performs similarly to the small-scale DFM.

We complement the modified Diebold-Mariano test by the fixed-b test of Coroneo and Iacone (2020) to assess the significance level for the predictive accuracy of the models. This test extends the traditional Diebold-Mariano test to the case of non-constant variance of the forecast error process and moreover, it extends the framework of Giacomini and White (2006) and Giacomini and Rossi (2010) by additionally controlling for small samples, which is particularly relevant to our application. The results of the fixed-b test of Coroneo and Iacone (2020) are provided in Table 6 for the NRMSE measure and in Table 16 of the appendix for the MAE measure. In the majority of pairwise comparisons, the results of the Coroneo and Iacone (2020) test confirm those of the Diebold-Mariano test in Tables 5 and 15.

Table 5: Diebold-Mariano test for GDP forecasts (2007-2018)

	Production	Expenditure	Income	TNT	Random walk	AR(1)	ARMA(2,1)	Small DFM	Large DFM
					3m(1q)				
Expenditure	<								
Income	>	>							
TNT	>	>	<						
Random walk	>***	>**	>*	>**					
AR(1)	>**	>**	>*	>*	<				
ARMA(2,1)	>***	>**	>	>	<	<			
Small DFM	>***	>*	<	>	<**	<	<		
Large DFM	>***	>***	>***	>***	>	>	>	>*	
MIDAS	>*	>	<	>	<**	<*	<	<	<**
					6m(2q)				
Expenditure	<								
Income	>*	>							
TNT	>	>	>						
Random walk	>***	>**	>***	>**					
AR(1)	>***	>*	>**	>	<				
ARMA(2,1)	>**	>	>**	>	<*	<**			
Small DFM	>***	>	>	>	<***	<	<		
Large DFM	>**	>	>	>	<**	<*	<	<	
MIDAS	>**	>	>	>	<**	<**	<	>	>
					9m(3q)				
Expenditure	>								
Income	>	>							
TNT	>	>*	>						
Random walk	>***	>**	>***	>*					
AR(1)	>**	>	>**	>	<**				
ARMA(2,1)	>*	>	>	>	<**	<***			
Small DFM	>	>	>	<	<***	<	<		
Large DFM	<	<	<	<	<***	<**	<*	<	
MIDAS	>	>	>	>	<**	<**	<	>	>*
					12m(4q)				
Expenditure	>								
Income	<	<							
TNT	>	>	>						
Random walk	>***	>**	>***	>**					
AR(1)	>**	>	>**	>	<***				
ARMA(2,1)	>**	>	>*	>	<***	<***			
Small DFM	>	>	>	>	<***	<	<		
Large DFM	<	<**	<	<**	<***	<**	<**	<	
MIDAS	>	>	>	<	<***	<***	<*	<	>**

The inequality sign compares a row against a column variable. The loss function is based on squared deviations.
The notation *** 1 percent; ** 5 percent; * 10 percent level of significance of a two-sided Diebold-Mariano test.

Table 6: Coroneo-Iacone test for GDP forecasts (2007-2018)

	Production	Expenditure	Income	TNT	Random walk	AR(1)	ARMA(2,1)	Small DFM	Large DFM
					3m(1q)				
Expenditure	$<$								
Income	$>$	$>$							
TNT	$>$	$>$	$<$						
Random walk	$>^{**}$	$>^{**}$	$>$	$>^{**}$					
AR(1)	$>^{**}$	$>^{**}$	$>^{**}$	$>$	$<$				
ARMA(2,1)	$>^{**}$	$>^{**}$	$>$	$>$	$<$	$<$			
Small DFM	$>^{**}$	$>^{**}$	$<$	$>$	$<^{**}$	$<$	$<$		
Large DFM	$>^{**}$	$>^{**}$	$>^{**}$	$>^{**}$	$>$	$>$	$>$	$>$	
MIDAS	$>$	$>$	$<$	$>$	$<^{**}$	$<^{**}$	$<^{**}$	$<^{**}$	$<^{**}$
					6m(2q)				
Expenditure	$<$								
Income	$>$	$>$							
TNT	$>$	$>$	$>$						
Random walk	$>^{**}$	$>^{**}$	$>^{**}$	$>$					
AR(1)	$>^{**}$	$>$	$>^{**}$	$>$	$<$				
ARMA(2,1)	$>^{**}$	$>$	$>^{**}$	$>$	$<$	$<$			
Small DFM	$>^{**}$	$>$	$>$	$>$	$<^{**}$	$<$	$<$		
Large DFM	$>^{**}$	$>$	$>$	$>$	$<$	$<$	$<$	$<$	
MIDAS	$>^{**}$	$>$	$>$	$>$	$<^{**}$	$<^{**}$	$<^{**}$	$>$	$>$
					9m(3q)				
Expenditure	$>$								
Income	$>$	$>$							
TNT	$>$	$>^{**}$	$>$						
Random walk	$>^{**}$	$>^{**}$	$>^{**}$	$>$					
AR(1)	$>^{**}$	$>$	$>^{**}$	$>$	$<$				
ARMA(2,1)	$>$	$>$	$>^{**}$	$>$	$<^{**}$	$<^{**}$			
Small DFM	$>$	$>$	$>$	$<$	$<^{**}$	$<$	$<$		
Large DFM	$<$	$<$	$<$	$<$	$<^{**}$	$<$	$<$	$<$	
MIDAS	$>$	$>$	$>$	$>$	$<^{**}$	$<^{**}$	$<$	$>$	$>$
					12m(4q)				
Expenditure	$>$								
Income	$<$	$<$							
TNT	$>$	$>$	$>$						
Random walk	$>^{**}$	$>^{**}$	$>^{**}$	$>^{**}$					
AR(1)	$>^{**}$	$>$	$>^{**}$	$>$	$<^{**}$				
ARMA(2,1)	$>^{**}$	$>$	$>$	$>$	$<^{**}$	$<^{**}$			
Small DFM	$>$	$>$	$>$	$>$	$<^{**}$	$<$	$<$		
Large DFM	$<$	$<$	$<$	$<^{**}$	$<^{**}$	$<^{**}$	$<$	$<$	
MIDAS	$>$	$>$	$>$	$<$	$<^{**}$	$<^{**}$	$<^{**}$	$<$	$>$

The inequality sign compares a row against a column variable. The loss function is based on squared deviations.
The notation ** 5 percent level of significance of a Coroneo-Iacone (fixed-b) test.

4.5 Predictive accuracy: Comparison with expert forecasts

Having compared the forecast precision of the CDFM to various competing models, we extend the comparison to expert forecasts. We follow Camacho and Pérez-Quirós (2010, 2011) and apply a forecast-encompassing test. This test is used to determine whether one of the forecasts encompasses all the relevant information from the other. It investigates whether there remains information in the CDFM's forecast error that can be explained by the professional forecasts. The resulting test statistics provide guidance on whether to combine different forecasts or discard a particular forecast that does not contain additional information.

To provide statistical evidence in terms of the predictive accuracy of the CDFM relative to expert forecasts, Table 7 presents the p-values of the forecast-encompassing test based on testing the significance of the parameter α_1 in the following OLS regression:

$$y_t - \hat{y}_t^{CDFM} = \alpha_0 + \alpha_1 \cdot \hat{y}_t^{PF} + \epsilon_t \tag{7}$$

where y_t is the realized GDP growth rate, and $\hat{y}_t^{CDFM}$ and $\hat{y}_t^{PF}$ are the real-time forecasts of the CDFM and of professional forecasters, respectively. To address the potential limitations of our short sample, we use robust standard errors as proposed by Bell and McCaffrey (2002).

The p-values indicate that the forecasts of the CDFM model encompass the forecasts from the professional forecasters. This holds across the CDFM's GDP forecast along the expenditure and income account, and the average GDP measure. As regards the GDP forecast from the production account, we find some evidence for improvement arising from the professional forecasts along the two-quarter horizon only. Despite this, we conclude that the CDFM's forecasts are hard to beat by any professional forecasts.

Table 7: Forecast encompassing test (p-values)

Forecast	3m(1q)	6m(2q)	9m(3q)	12m(4q)
		GDP production		
EC	0.374	0.059	0.188	0.700
IMF	0.442	0.024	0.160	0.500
OECD	0.287	0.022	0.072	0.754
WIFO	0.341	0.031	0.144	0.862
OeNB	0.413	0.009	0.110	0.886
		GDP expenditure		
EC	0.489	0.489	0.621	0.517
IMF	0.360	0.360	0.570	0.703
OECD	0.405	0.405	0.415	0.301
WIFO	0.552	0.552	0.578	0.369
OeNB	0.542	0.542	0.447	0.472
		GDP income		
EC	0.325	0.295	0.325	0.828
IMF	0.508	0.353	0.508	0.899
OECD	0.182	0.301	0.182	0.782
WIFO	0.321	0.400	0.321	0.583
OeNB	0.223	0.392	0.223	0.450
		GDP average		
EC	0.674	0.160	0.353	0.674
IMF	0.657	0.143	0.397	0.657
OECD	0.516	0.105	0.191	0.516
WIFO	0.739	0.267	0.375	0.739
OeNB	0.889	0.224	0.243	0.889

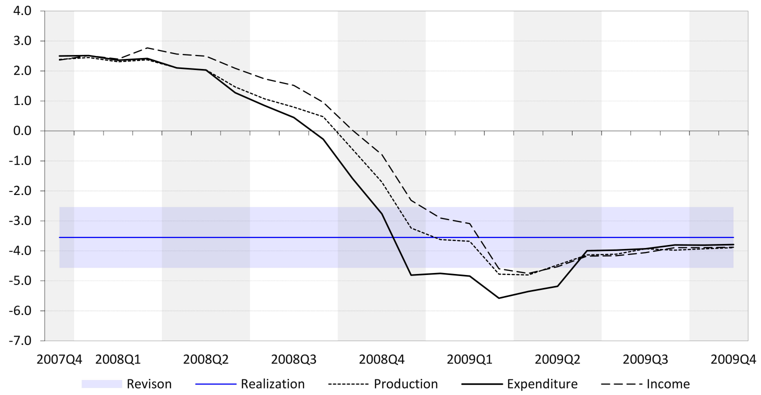
The p-values refer to the results of a t-test on the parameter α_1 of equation (7). The acronyms refer to European Commission (EC), International Monetary Fund (IMF), Organisation for Economic Co-operation and Development (OECD), Austrian Institute of Economic Research (WIFO), and Austrian National Bank (OeNB).

4.6 Forecast trajectory

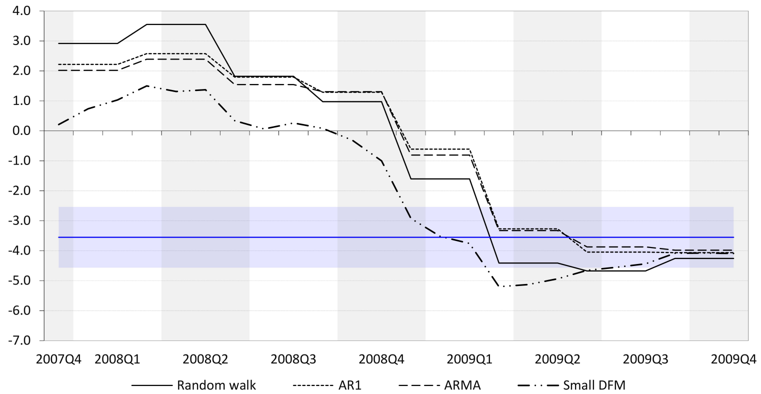
In addition to investigating the CDFM's forecasting performance over a long sample, we also assess the model's performance during specific historic episodes. In this context, we can assess the role of new, updated incoming information for the forecast. Our (pseudo) real-time data-set allows for such an assessment. We focus our analysis on the global financial crisis of 2008/2009.

We compute real-time forecasts for the four quarters of 2009 to establish the annual growth rate of this year. We carry out this exercise with the information available at different points in time starting in January 2008. The path of the forecast trajectory is displayed in subplot (a) in Figure 3.

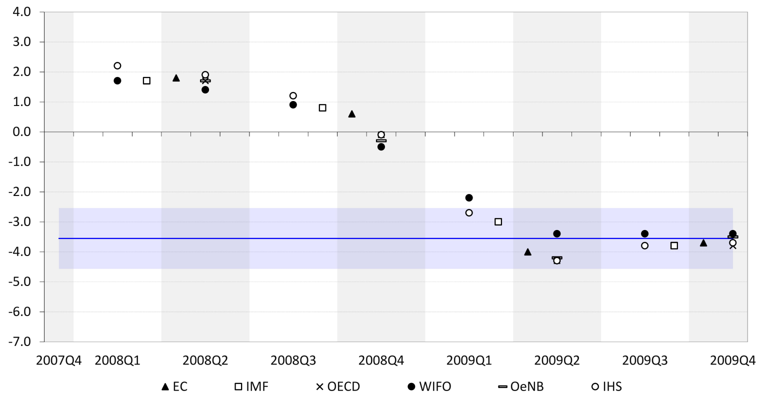
Figure 3: Forecasting performance 2009



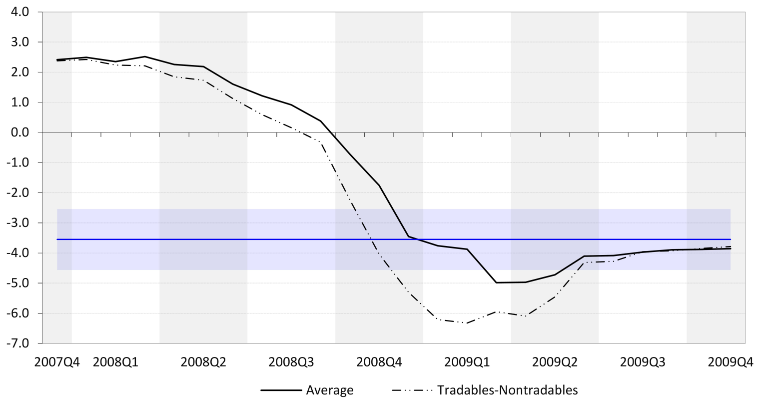
(a) Cluster of Dynamic Factor Models (CDFM)



(b) Competing models



(c) Expert forecasts



(d) CDFM with TNT extension

Note: The figure plots real-time forecasts of the CDFM (in sub-plot (a) and (d)), the competing models (sub-plot(b)) and expert forecasts (sub-plot(c)) for the annual GDP growth rate for the year 2009 established during 2008 and 2009. The values are in percent. The acronyms for the expert forecasts refer to European Commission (EC), International Monetary Fund (IMF), Organization for Economic Co-operation and Development (OECD), Austrian Institute of Economic Research (WIFO), Institute for Advanced Studies (IHS) and Austrian National Bank (OeNB).

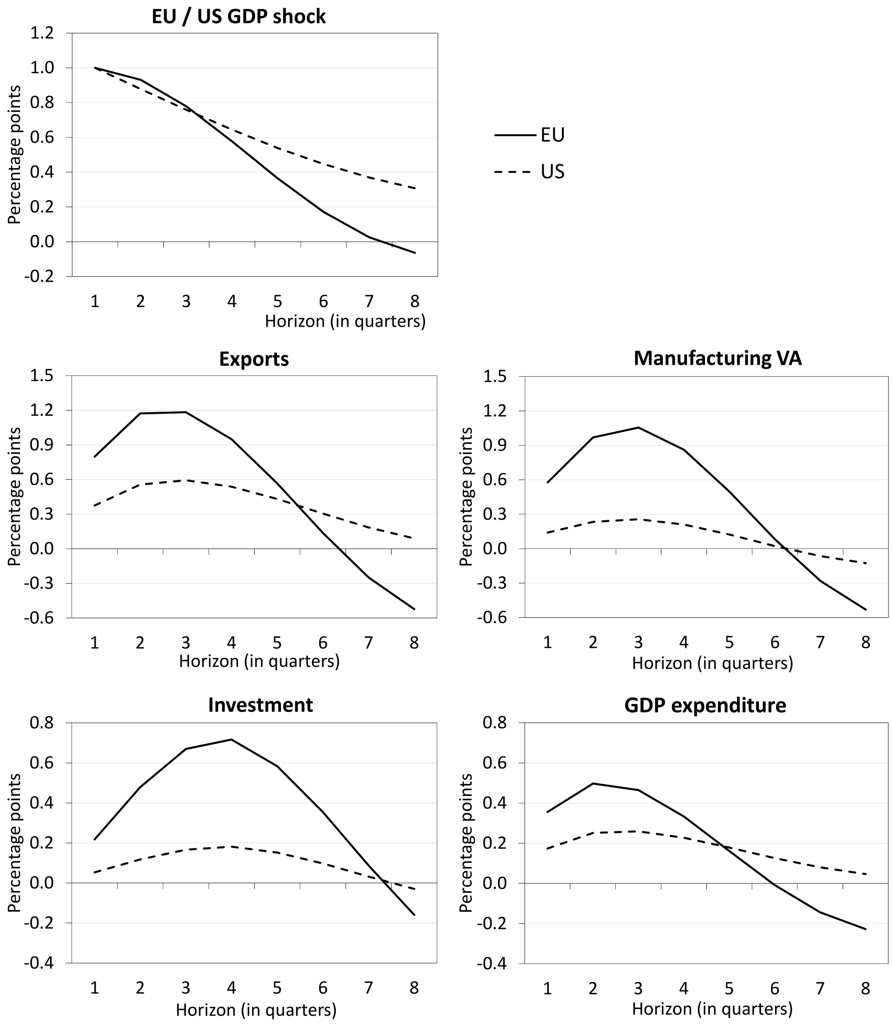
This figure helps us to address a question that has been the source of intensive debates in many countries: When did the authorities realize that the downturn had started? It is worth recalling that forecasting this turning point was a rather difficult task. The financial turmoil had increased the forecast uncertainty to unprecedented levels. In addition, at the beginning of the recession period, the financial variables and soft indicators were giving signals of a recession that were not associated with clear signals from real activity. Finally, for many countries it turned out to be the first negative annual growth after a long period of sustained growth. Figure 3 shows that signals of a business cycle turning point started to become clear around the summer of 2008.

For comparison, forecasts from the competing models and the expert forecasts (IMF, OECD, EC, etc.) are shown in subplot (b) and (c). Each subplot is extended for the actual GDP growth values (blue solid line) jointly with a measure of average revisions of GDP growth (blue band, as in Figure 2). This figure shows several noticeable features which illustrate the advantages of real-time forecasting with the CDFM against alternative approaches. All forecasts display a declining path. However, the CDFM's forecasts display the quickest downward adjustment. The CDFM anticipated negative growth rates for 2009 from August/September 2008 onwards, while most of the competing models' forecasts did so only from the end of 2008. The small-scale DFM for GDP is comparable to the CDFM in terms of detecting the point in time when growth for 2009 turned negative. The same applies to the large-scale DFM and the MIDAS regression model (not shown in the figure).

Regarding the expert forecasts, it should be noted that the hesitant adjustment of the annual forecast in the form of excessive restraints as regards the publication of negative growth rates weighs heavily on the forecast error of the expert forecasts. This appears to be particularly pronounced at the end of the year. The CDFM forecast was already at -3.6 percent (average across the three accounts) and thus already very close to the realized value (-3.5 percent); the expert forecasts, on the other hand, were only slightly negative (-0.5 percent) and in some cases even positive.

Although a reluctance to publish negative growth rates may also result from the ambition not to spread excessive pessimism, the credibility of the expert forecasts is likely to suffer increasingly if such forecast errors occur repeatedly. This systematic bias does not apply to a model forecast though. The forecast is unbiased, objective and can be illustrated easily and transparently by means of the indicators used.

Figure 4: Increase in foreign GDP growth



Note: The effect of a 1 percentage point increase in the EU or US real GDP growth rate.

4.7 Scenario analysis: an increase in foreign GDP growth

We perform a scenario analysis as suggested by Bańbura et al. (2015) to show the functionality of the CDFM beyond pure forecasting. The following extends the ideas put forth in Section 3.4 as regards the working of the CDFM. We evaluate the effects of two scenarios: Scenario (i) higher GDP growth in the US; Scenario (ii) higher GDP growth in the EU. To this purpose, we estimate the CDFM using the entire sample and produce two forecasts: an unconditional forecast for $T+1,\ldots,T+h$ given the sample $1,2,\ldots,T$ ('baseline'), and two conditional forecasts in which the GDP growth rate in (i) the US, and (ii) the EU in $T+1$ are set to the value of its own unconditional forecast plus 1 percentage point ('shock'), with all the remaining variables left unconstrained. The scenario results for the variables are computed by taking the difference of conditional and unconditional forecasts. This is equivalent to computing a generalized impulse response function to an increase in the GDP growth rate of the US or the EU, as highlighted in Bańbura et al. (2015). We consider a horizon of eight-quarters ($h = 8$).

The responses shown in Figure 4 are limited to the variables that align with the discussion in Section 3.4.[16] All results are reported in terms of differences of the year-over-year (yoy) growth rates of the variables in the shock scenario compared to the baseline. In each case, real GDP for one of the two foreign economies is one percentage point higher on impact. The domestic real economy (GDP, exports, manufacturing value added, and investment) closely mirrors the path of foreign GDP growth. The foreign shock enters the domestic economy through exports of goods and services, which rise on impact, continue to rise for three quarters, and then converge back to the levels prevailing before the shock (uniformly in the case of the EU GDP growth shock and in a hump-shaped pattern in the case of the US GDP growth shock). Exports, in turn, boost value added in manufacturing, whose response has a similar shape but a smaller magnitude to that of exports. The increase in manufacturing activity induces an increase in investment, especially investment in equipment and construction. Despite the positive effect, the reaction in investment is rather sluggish, suggesting a lagged investment response, which could theoretically be motivated by a time-to-build effect (Majd and Pindyck, 1987; Zhou, 2000).

The results also suggest that aggregate shocks originating in the EU tend to have a larger effect than the US counterparts.[17] The simulations only show the direct effects of foreign trade shocks, however, it is very likely that an expansionary shock emanating from the US will not only affect the domestic economy directly, but also indirectly, e.g. via the EU economy. While these indirect effects are ignored in the current setup, as they are not the focus of the analysis, they could easily be addressed by expanding the behavioral DFMs of the external sector.

[16] Rather than reporting the whole distribution of the responses for each variable as in Bańbura et al. (2015), we only show the point estimates. This is due to the fact that we do not use a Bayesian approach in the CDFM for inference. Putting the CDFM into a Bayesian framework not only allows for an easy computation of error bands for the difference of the conditional and unconditional forecasts, but also to take the forecast uncertainty of downstream models into account for the forecast uncertainty of GDP.

[17] This is mainly due to the fact that the Austrian economy has a higher foreign trade intensity with EU countries, accounting for around 75 percent of Austrian exports, than with the US.

4.8 Limitations of the CDFM

The discussion on the working of the CDFM as outlined in Sections 3.4 and 4.7 suggests some limitations. The main technical limitation is the sequential nature of the CDFM as it precludes simultaneous determination of forecasts. This means that a forecast for a downstream variable does not affect the forecast for an upstream variable. This underscores the importance of determining an optimal sequence that aligns with economic theory and optimizes forecast accuracy, as achieved by the Granger causality tests. The path dependence of the CDFM forecast and the consequent lack of feedback effects limits the use of the CDFM[18] for nowcasting and short-term forecasting as longer projections may require closed feedback loops. This is one of the key results of the forecasting assessment of Section 4.2. In principle feedback effects could be allowed for in the CDFM. This would, however, require to introduce a balancing procedure, as suggested in van der Ploeg (1982).

5 Summary

In this paper, we propose a methodology for large-scale macroeconomic nowcasting and forecasting using a rich set of economic indicators available on a monthly basis. To this end, we use the concept of unidirectional Granger-causality to link a series of small-scale dynamic factors models in a cluster. Since the individual models are estimated using the Kalman filter, the resulting *Cluster of Dynamic Factor Models* can handle data at different frequencies, as well as data that feature gaps and ragged edges resulting from asynchronous publication.

We find that the nowcasting and forecasting performance of the CDFM is superior to that of naive forecasting models and purely technical time series models that allow for rich dynamics. Moreover, the CDFM's forecasts for GDP are noticeably more accurate than those of classical small-scale dynamic factor models directly applied to GDP. Finally, the forecasts produced by the CDFM compete well with professional forecasts.

While the CDFM provides a transparent and highly flexible structure, it confers several further advantages. First, each individual DFM from the cluster can be used indepen-

[18] This suggests that the structure of the CDFM can be considered as a restricted large-scale factor model, in which the restrictions pertain to *zero-restrictions* on specific factor loadings.

dently. This allows for producing both unconditional forecasts when used individually and conditional forecasts when used within the CDFM. Second, the CDFM allows forecasters to introduce their own judgment, which is integrated into a consistent conditional forecast. Third, the CDFM allows for scenario analysis in the context of forecasting on a large-scale, which renders our approach particularly useful for assessing the sensitivity of a macroeconomic forecast with respect to the underlying assumptions. Fourth, the CDFM not only allows to forecast real-time GDP, but also to incorporate information on the components of the expenditure, production and income accounts that determine the GDP forecast, providing broader insight into the causes of revisions in GDP forecasts. Finally, the CDFM has a simple structure that is easy to replicate. This makes the proposed modeling approach particularly attractive for applied work.

To summarize, we believe that the *Cluster of Dynamic Factor Models* presented in this paper is a practical tool for nowcasting and short-term forecasting on a large-scale. It has a good forecasting record, is automatically updated when new information becomes available, provides a way of measuring the effects of new developments in GDP indicators and their subcomponents, and allows extensions to be implemented quickly due to its transparent and simple framework.

Declarations

Data availability statement.

The data used in this study are available from the authors upon request.

Conflict of interest.

The authors declare that they have no conflict of interest.

Ethical approval.

This article does not contain any studies with human participants or animals performed by any of the authors.

References

Aastveit, Knut, and Tørres Trovik (2012) 'Nowcasting Norwegian GDP: The role of asset prices in a small open economy.' *Empirical Economics* 42(1), 95–119

Angelini, Elena, Gonzalo Camba-Mendez, Domenico Giannone, Lucrezia Reichlin, and Gerhard Rünstler (2011) 'Short-term forecasts of euro area GDP growth.' *Econometrics Journal* 14(1), 25–44

Arnoštová, Kateřina, David Havrlant, Luboš Růžička, and Peter Tóth (2011) 'Short-term forecasting of Czech quarterly GDP using monthly indicators.' *Czech Journal of Economics and Finance* 61(6), 566–583

Bai, Jennie, Eric Ghysels, and Jonathan H. Wright (2013) 'State space models and midas regressions.' *Econometric Reviews* 32(7), 779–813

Bai, Jushan, and Peng Wang (2016) 'Econometric analysis of large factor models.' *Annual Review of Economics* 8(1), 53–80

Bai, Jushan, and Serena Ng (2002) 'Determining the number of factors in approximate factor models.' *Econometrica* 70(1), 191–221

Banbura, Marta, and Gerhard Rünstler (2011) 'A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP.' *International Journal of Forecasting* 27(2), 333–346

Bańbura, Marta, Domenico Giannone, and Michele Lenza (2015) 'Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections.' *International Journal of Forecasting* 31(3), 739–756

Barhoumi, Karim, Gerhard Rünstler, Riccardo Cristadoro, Ard Den Reijer, Audrone Jakaitiene, Piotr Jelonek, António Rua, Karsten Ruth, Szilard Benk, and Christophe Van Nieuwenhuyze (2008) 'Short-term forecasting of GDP using large monthly datasets: A pseudo real-time forecast evaluation exercise.' Working papers 215, Banque de France, July

Bell, Robert M., and Daniel McCaffrey (2002) 'Bias reduction in standard errors for linear regression with multi-stage samples.' *Survey Methodology* 28, 169–181

Belloni, A., V. Chernozhukov, and K. Kato (2014) 'Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems.' *Biometrika* 102(1), 77–94

Boivin, Jean, and Serena Ng (2006) 'Are more data always better for factor analysis?' *Journal of Econometrics* 132(1), 169–194

Bårdsen, Gunnar, Ard den Reijer, Patrik Jonasson, and Ragnar Nymoen (2012) 'MOSES: Model for studying the economy of Sweden.' *Economic Modelling* 29(6), 2566–2582

Brayton, Flint, Andrew Levin, Ralph Lyon, and John C. Williams (1997) 'The evolution of macro models at the Federal Reserve Board.' *Carnegie-Rochester Conference Series on Public Policy* 47, 43–81

Camacho, Maximo, and Agustín García-Serrador (2014) 'The Euro-Sting revisited: The usefulness of financial indicators to obtain Euro Area GDP forecasts.' *Journal of Forecasting* 33(3), 186–197

Camacho, Maximo, and Gabriel Pérez-Quirós (2010) 'Introducing the euro-sting: Short-term indicator of euro area growth.' *Journal of Applied Econometrics* 25(4), 663–694

—— (2011) 'Spain-sting: Spain short-term indicator of growth.' *Manchester School* 79(s1), 594–616

Clark, Todd (2004) 'Can out-of-sample forecast comparisons help prevent overfitting?' *Journal of Forecasting* 23(2), 115–139

Clements, Michael, and David Hendry (1998) *Forecasting Economic Time Series* (Cambridge University Press)

Cobb, Marcus P. A. (2020) 'Aggregate density forecasting from disaggregate components using Bayesian VARs.' *Empirical Economics* 58(1), 287–312

Coroneo, Laura, and Fabrizio Iacone (2020) 'Comparing predictive accuracy in small samples using fixed-smoothing asymptotics.' *Journal of Applied Econometrics* 35(4), 391–409

Dias, Francisco, Maximiano Pinheiro, and António Rua (2015) 'Forecasting Portuguese GDP with factor models: Pre- and post-crisis evidence.' *Economic Modelling* 44, 266–272

Diebold, Francis X. (1998) 'The past, present, and future of macroeconomic forecasting.' *Journal of Economic Perspectives* 12(2), 175–192

Diebold, Francis X., and Roberto S. Mariano (1995) 'Comparing predictive accuracy.' *Journal of Business & Economic Statistics* 13(3), 253–263

Doz, Catherine, Domenico Giannone, and Lucrezia Reichlin (2011) 'A two-step estimator for large approximate dynamic factor models based on Kalman filtering.' *Journal of Econometrics* 164(1), 188–205

Durbin, James, and Siem Jan Koopman (2001) *Time Series Analysis by State Space Methods* (Oxford University Press)

Eitrheim, Øyvind, Tore Anders Husebø, and Ragnar Nymoen (1999) 'Equilibrium-correction vs. differencing in macroeconometric forecasting.' *Economic Modelling* 16(4), 515–544

Esteves, Paulo Soares (2013) 'Direct vs bottom–up approach when forecasting GDP: Reconciling literature results with institutional practice.' *Economic Modelling* 33, 416–420

Eurostat (2013) *Handbook on quarterly national accounts* (Luxembourg: Publications Office of the European Union)

Forni, Mario, and Marco Lippi (2001) 'The Generalized Dynamic Factor Model: Representation Theory.' *Econometric Theory* 17(6), 1113–1141

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin (2000) 'The Generalized Dynamic-Factor Model: Identification And Estimation.' *The Review of Economics and Statistics* 82(4), 540–554

Foroni, Claudia, and Massimiliano Marcellino (2014) 'A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates.' *International Journal of Forecasting* 30(3), 554–568

Friesenbichler, Klaus S., and Christian Glocker (2019) 'Tradability and productivity growth differentials across EU Member States.' *Structural Change and Economic Dynamics* 50, 1–13

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov (2006) 'Predicting volatility: getting the most out of return data sampled at different frequencies.' *Journal of Econometrics* 131(1), 59–95

Giacomini, Raffaella, and Barbara Rossi (2010) 'Forecast comparisons in unstable environments.' *Journal of Applied Econometrics* 25(4), 595–620

Giacomini, Raffaella, and Halbert White (2006) 'Tests of conditional predictive ability.' *Econometrica* 74(6), 1545–1578

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

Glocker, Christian, and Philipp Wegmüller (2020) 'Business cycle dating and forecasting with real-time Swiss GDP data.' *Empirical Economics* 58(1), 73–105

Glocker, Christian, and Serguei Kaniovski (2014) 'A financial market stress indicator for Austria.' *Empirica - Journal of European Economics* 41(3), 481–504

Glocker, Christian, and Werner Hölzl (2021) 'A direct measure of subjective business uncertainty.' *German Economic Review (forthcoming)*

Granger, Clive, and Yongil Jeon (2004) 'Forecasting performance of information criteria with many macro series.' *Journal of Applied Statistics* 31(10), 1227–1240

Hamilton, James D. (1994) *Time Series Analysis* (Princeton University Press)

Hammersland, Roger, and Cathrine Bolstad Træ(2014) 'The financial accelerator and the real economy: A small macroeconometric model for Norway with financial frictions.' *Economic Modelling* 36, 517–537

Harvey, David, Stephen Leybourne, and Paul Newbold (1997) 'Testing the equality of prediction mean squared errors.' *International Journal of Forecasting* 13(2), 281–291

Hecq, Alain, Luca Margaritella, and Stephan Smeekes (2019) 'Granger Causality Testing in High-Dimensional VARs: a Post-Double-Selection Procedure.' Papers 1902.10991, arXiv.org, February

Heilemann, Ullrich, and Hagen Findeis (2012) 'Empirical determination of aggregate demand and supply curves: The example of the RWI Business Cycle Model.' *Economic Modelling* 29(2), 158–165

Heinisch, Katja, and Rolf Scheufele (2018) 'Bottom-up or direct? Forecasting German GDP in a data-rich environment.' *Empirical Economics* 54(2), 705–745

Hmamouche, Youssef (2020) 'NlinTS: An R Package For Causality Detection in Time Series.' *The R Journal* 12(1), 21–31

Jiang, Yu, Yongji Guo, and Yihao Zhang (2017) 'Forecasting China's GDP growth using dynamic factors and mixed-frequency data.' *Economic Modelling* 66, 132–138

Kapetanios, George, Vincent Labhard, and Simon Price (2008) 'Forecast combination and the Bank of England's suite of statistical forecasting models.' *Economic Modelling* 25(4), 772–792

Klein, Lawrence R., and Süleyman Özmucur (2010) 'The use of consumer and business surveys in forecasting.' *Economic Modelling* 27(6), 1453–1462

Kuck, Konstantin, and Karsten Schweikert (2021) 'Forecasting Baden-Württemberg's GDP growth: MIDAS regressions versus dynamic mixed-frequency factor models.' *Journal of Forecasting* 40(5), 861–882

Kuzin, Vladimir, Massimiliano Marcellino, and Christian Schumacher (2013) 'Pooling versus model selection for nowcasting with many predictors: empirical evidence for six industrialized countries.' *Journal of Applied Econometrics* 28(3), 392–411

Leeb, Hannes, and Benedikt M. Pötscher (2005) 'Model selection and inference: Facts and fiction.' *Econometric Theory* 21(1), 21–59

Lehmann, Robert (2021) 'Forecasting exports across Europe: What are the superior survey indicators?' *Empirical Economics* 60(5), 2429–2453

Liu, Philip, Troy Matheson, and Rafael Romeu (2012) 'Real-time forecasts of economic activity for Latin American economies.' *Economic Modelling* 29(4), 1090–1098

Majd, Saman, and Robert S. Pindyck (1987) 'Time to build, option value, and investment decisions.' *Journal of Financial Economics* 18(1), 7–27

Marcellino, Massimiliano, and Christian Schumacher (2010) 'Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP.' *Oxford Bulletin of Economics and Statistics* 72(4), 518–550

Marcellino, Massimiliano, and Vasja Sivec (2021) 'Nowcasting GDP growth in a small open economy.' *National Institute Economic Review* 256, 127–161

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

Modugno, Michele, Barış Soybilgen, and Ege Yazgan (2016) 'Nowcasting Turkish GDP and news decomposition.' *International Journal of Forecasting* 32(4), 1369–1384

Moser, Gabriel, Fabio Rumler, and Johann Scharler (2007) 'Forecasting Austrian inflation.' *Economic Modelling* 24(3), 470–480

Öğünç, Fethi, Kurmaş Akdoğan, Selen Başer, Meltem Gülenay Chadwick, Dilara Ertuğ, Timur Hülagü, Sevim Kösem, Mustafa Utku Özmen, and Necati Tekatlı(2013) 'Short-term inflation forecasting models for Turkey and a forecast combination analysis.' *Economic Modelling* 33, 312–325

Pareja, Ana Arencibia, Ana Gómez-Loscos, Mercedes de Luis López, and Gabriel Pérez-Quirós (2020) 'A short-term forecasting model for the Spanish economy: GDP and its demand components.' *Economía* 43(85), 1–30

Proietti, Tommaso, Alessandro Giovannelli, Ottavio Ricchi, Ambra Citton, Christían Tegami, and Cristina Tinti (2021) 'Nowcasting GDP and its components in a data-rich environment: The merits of the indirect approach.' *International Journal of Forecasting, In Press*

Qin, Duo, Marie Anne Cagas, Geoffrey Ducanes, Xinhua He, Rui Liu, Shiguo Liu, Nedelyn Magtibay-Ramos, and Pilipinas Quising (2007) 'A macroeconometric model of the Chinese economy.' *Economic Modelling* 24(5), 814–822

Rusnák, Marek (2016) 'Nowcasting Czech GDP in real time.' *Economic Modelling* 54, 26–39

Schumacher, Christian, and Jörg Breitung (2008) 'Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data.' *International Journal of Forecasting* 24(3), 386–398

Stock, James H., and Mark W. Watson (1991) 'A probability model of the coincident economic indicators.' In *Leading Economic Indicators: New Approaches and Forecasting Records,* ed. Kajal Lahiri and Geoffrey H. Moore (Cambridge, UK: Cambridge University Press) pp. 63–90

van der Ploeg, Frederick (1982) 'Reliability and the adjustment of sequences of large economic accounting matrices.' *Journal of the Royal Statistical Society* 145(2), 169–194

Wladyslaw, Welfe (2011) 'Long-term macroeconometric models: The case of Poland.' *Economic Modelling* 28(1), 741–753

Zhou, Chunsheng (2000) 'Time-to-build and investment.' *Review of Economics and Statistics* 82(2), 273–282

A Alternative Granger causality tests

We check the links uncovered by the conventional Granger causality test using two alternative tests. The Granger is a bivariate test that cannot account for the possibility of a causal relationship between a pair of variables being induced by a third variable acting as a common cause. The omission of common causes can lead to spurious causality. The second issue is that the Granger test involves model selection as a first step, which invalidates the subsequent asymptotic and finite sample inference (Leeb and Pötscher, 2005).

The generalization by Hecq et al. (2019) rectifies the deficiency of a bivariate causality test. This multivariate test is based on a high-dimensional VAR that includes the entire dataset used to estimate the CDFM. The estimation procedure uses a sparsity-seeking regularization that uncovers the key dynamic interactions among the variables while discarding the rest. Model selection and testing follow the Belloni et al. (2014) post-double-selection approach that partially mitigates the problem associated with the validity of post-model-selection estimators. The LM test then accounts for multiple joint cause for each pair of variables of interest. We use test statistic with a final sample correction that has been shown to improve the size of the test.

The conventional Granger causality test and its multivariate generalization involve a VAR and thus assume a linear relationship between the variables. Our second alternative is based on a highly nonlinear View Adaptive Recurrent Neural Network (VA-RNN) model, which is a multilayer feed-forward neural network (Hmamouche, 2020). The nonlinear test is bivariate. The test statistic being equivalent to that of the conventional Granger test in that it relates the residual sum of squares of errors of the restricted and the unrestricted models and follows an F-distribution under the null hypothesis of no causality.

Table 8 summarizes the test results for the final selection of links. The underlying models have been tested using quarterly data up to lag two, as suggested by the BIC criterion. Due to a sequential structure of the cluster we are only interested in unidirectional causality relationships, by which past realizations of the variable listed in the first column improves forecasts of the corresponding variable in the second column, but not the other way around. We call unidirectional causality weak (W) if the test is significant at the 10 percent level, and strong (S) if it is significant at the 5 percent level.

Note that the multivariate test tends to uncover fewer causal links that the two bivariate tests, which is expected due to the existence of multiple common causes in large dataset of highly interdependent time series. All the links except those between the exports of services and the value added of the service sector, and between construction investment and the value added in the construction sector have been validated by at least one of the three tests. These two links have been retained because their inclusion has significantly improved the forecasting performance of the respective value-added models when entered *contemporaneously*. Contemporaneous links could not be validated by causality tests that seek an intertemporal relationship between a pair of variables.

Table 8: Alternative causality tests

From	To	Classic	Mult. Lag 1	Nonlinear	Classic	Mult. Lag 2	Nonlinear
Exports of goods	Investment intangibles	S	S	S	S		W
Exports of goods	Manufacturing VA	S			S	S	
Exports of goods	Capital income	S					
Exports of services	Services VA						
Investment equipment	Investment intangibles				S	S	
Investment construction	Construction VA						
Manufacturing VA	Investment equipmen	S		S	S		S
Manufacturing VA	Investment construction	S	S	S	S	S	W
Manufacturing VA	Construction VA	W	S		W		
Manufacturing VA	Services VA	S		S	S		S
Manufacturing VA	Labor income			S			S
Construction VA	Labor income	S		S			S
Services VA	Investment equipment	S		W	S		
Services VA	Labor income			S			S
Labor income	Consumption		S				
Capital income	Investment equipment	S		S	S		W
Capital income	Consumption					S	

(W) denote unidirectional causality links significant at the 10 percent level, whereas (S) denote links significant at the 5 percent level.

B An alternative view on the production account: tradable and nontradable goods (TNT)

The approach proposed in Section 2 for modeling production is frequently used by international organizations, central banks and economic research institutes. It follows the sectoral composition of the National Accounts, but does not reflect the macroeconomic theory, which usually groups the sectors in producers of tradable goods (T) and non-tradable goods (NT).[19] While the TNT classification is clear for some sectors, it can be ambiguous for others. Moreover, structural change might turn a previously tradable sector into a nontradable one, and vice versa. We follow the approach to the sectoral classification in Friesenbichler and Glocker (2019), determine the nominal value added for

[19] We use the term *goods* for both goods and services.

tradable goods and nontradable goods and calculate the corresponding deflators, which allows us to determine the real value added of these two categories.

The next step involves specification of separate behavioral models for tradable goods and nontradable goods, and an aggregator model for the GDP. The aggregator model features an error term, as the tradable and nontradable goods do not sum up to GDP, the difference being product taxes and subsidies. Another reason for including an error term is that we again consider a log-linearized representation of a weighted sum, in which the weights of the components can change over time. These changes in the weights are addressed by an autoregressive error term, as shown in equation (5).

Table 9: Tradables (T) and nontradables (NT) (2007-2018)

Variable	3m(1q)	6m(2q)	9m(3q)	12m(4q)
Tradables VA	0.44	0.65	0.92	1.08
Nontradables VA	0.47	0.69	0.86	0.99
GDP average	0.41	0.61	0.86	1.05
GDP (TNT)	0.44	0.71	0.98	1.12
GDP Small DFM	0.50	0.76	0.96	1.14
Tradables deflator	0.64	0.89	1.05	1.10
Nontradables deflator	0.38	0.55	0.68	0.78
GDP deflator (TNT)	0.64	0.83	0.91	0.94
GDP deflator	0.57	0.73	0.76	0.76

Table 3 provides the normalized values of the RMSE for a GDP forecast based on the TNT approach. The forecasting accuracy of the TNT approach is similar to that of the standard production-side GDP approach considered in Section 3. Although the normalized RMSE of GDP is slightly smaller than in the conventional three-sector approach (manufacturing, construction, services), this difference is not statistically significant. The forecasts of the value added of tradable goods and nontradable goods seem to be comparatively precise. The values for the normalized RMSE for these two components are smaller than those for the sectors in the conventional approach for all forecasting horizons.

Subplot (d) in Figure 3 shows the annual GDP-forecast for the year 2009 obtained using the TNT approach over a period of 2008 and 2009. The TNT-based forecasts indicate negative annual growth relatively early, somewhat overestimating the extent of the recession at the end of 2008, but nonetheless approaching the realized value quickly.

Finally, the deflators for tradable and nontradable goods allow for an alternative approach to modeling and forecasting the GDP deflator. We specify a behavioral model for each of the two deflators and link them to the GDP deflator in an aggregator model. The forecast evaluation in Table 9 shows the normalized RMSE of the deflators for tradable and nontradable goods to be comparatively small. The TNT-based forecasts of the GDP deflator are thus comparable to the baseline approach.

C The small-scale dynamic factor model

We consider a small-dynamic factor model (*small DFM*), as popularized by Mariano and Murasawa (2003); Camacho and Pérez-Quirós (2010, 2011); Arnoštová et al. (2011); Aastveit and Trovik (2012) as a competing model. This approach comprises a small-scale and hence simple factor model applied directly to GDP growth. Following Mariano and Murasawa (2003), we combine monthly and quarterly data, expressing the quarterly data as a function of monthly data. If the sample mean of the three monthly observations in a given quarter can be approximated by the geometric mean, then the quarterly growth rates can be decomposed as weighted averages of monthly growth rates. We follow the outline put forward in Section 3 and utilize the approach motivated by Glocker and Wegmüller (2020) to select an appropriate set of variables. This approach explicitly takes into account the fact that additional variables do not necessarily improve the model's forecast. The available set of variables contains around sixty variables.

The principal criterion for variable selection is out-of-sample forecasting ability, producing a set of variables geared towards economic expectations. We have already seen in Section 4.6 that the resulting model performs well in forecasting the 2009 economic downturn. The final specification of the small-scale factor model (SDFM) for GDP includes: (1) expectations in the construction sector, (2) expectations in the manufacturing sector, (3) expectations in the service sector, (4) Purchasing Managers Index (PMI), (6) order backlog (manufacturing sector), (7) employment (all sectors), (8) vacancies (all sectors), (9) retail sales (total) and (10) truck mileage. We add further variables only if they improve the out-of-sample forecasting performance of the model. We find that some additional variables could be included, however, they do not improve the forecast (e.g, Economic Sentiment Index (ESI) from the European Commission, ATX/Austrian Traded Index volatility, the financial market stress indicator as considered in Glocker and Kaniovski (2014), term-structure – i.e. the difference between 10-year and 2-year government bond yield, industrial production – excluding the construction sector, and retail sales). Other variables worsened the out-of-sample forecasts and were subsequently discarded from the model. The final selection proved robust to enlargements of the model in various directions. We tested our model using disaggregated versions of the variables already included in the model. For instance, we used retail sales without oil-related products instead of total retail sales. We also checked for the employment of different sectors (manufacturing sector, construction sector) instead of the aggregate measure, failing to improve the model in all cases.

Since the variables considered address both the outlook and the current situation, we allow for a temporal displacement between GDP as the target variable and the additional variables, for which we follow Camacho and García-Serrador (2014). This set-up follows Camacho and Pérez-Quirós (2010) with a dynamic factor structure involving one factor with two lags (we omit the elements concerning data revisions from the model). The number of factors is selected by using Bai and Ng (2002) information criteria (BG) modified to take into account that the parameters are estimated using maximum likelihood. The number of lags in the factor equation and for the error terms was chosen by relying on the Bayesian Information Criterion (BIC).

D The large-scale dynamic factor model

In addition to the small-scale dynamic factor model as competing model, we also consider a large-scale dynamic factor model (*large DFM*). Large factor models use a small number of factors to capture the co-movement of a high-dimensional set of time-series. High dimensionality poses challenges. The first concerns the data frequency and missing observations in general. The second concerns the identification of the latent factors.

In order to allow for a decent comparison of the predictive accuracy of the CDFM to the large DFM, we use all variables of the CDFM in the large DFM. We extract the latent factors by relying on principal component analysis. This approach is easy to implement, and, given that the cross-section and time dimension are large, provides consistent estimates under quite general assumptions. It suffers, however, from one main drawback: the dataset must be balanced, that is, the start and end points have to be the same across all observable variables and all data-series must have the same frequency so that missing observations do not arise. To this purpose, we consider a quarterly frequency of the data and of the large DFM alike. We transform all monthly series into quarterly series by considering a three-months average. The sample used for the estimation starts in 2007 as missing observations prior to this year would otherwise impede the estimation. All series enter the large DFM contemporaneously, hence, in contrast to the small DFM, we do not allow for temporal displacements within the data series. We rely on Bai and Ng (2002) to determine the number of factors, and specify a finite-order VAR model to approximate the dynamics of the latent factors.

We consider the principal component methods and maximum likelihood methods to estimate the large DFM (see Bai and Wang, 2016, for further details). To this purpose, we again standardize all data-series prior to the estimation. The estimation relies on a two-step procedure in which the latent factors are estimated in a first step, and the VAR model in the second step. As highlighted by Doz et al. (2011), this procedure yields consistent estimates even when the static factor model is misspecified with respect to some of its dynamic elements. We establish forecasts from the large DFM in the same form as done for the other models. This allows for an adequate comparison of the large DFM's forecasts with those of the other models, keeping in mind, though, the different underlying frequency.

E The mixed-data-sampling (MIDAS) regression model

We complete the set of competing models with a mixed-data-sampling (*MIDAS*) regression model for GDP using the set of indicators from the small-scale dynamic factor model detailed in Section C. MIDAS regression was developed by Ghysels et al. (2006) as a means of predicting a single low-frequency time series (quarterly GDP growth) with multiple high-frequency indicators (monthly indicators). The key feature of MIDAS regression is the distribution of the current and past values of the high-frequency indicators that effectively yields different forecasting models for each forecast horizon. We found that a MIDAS specification with simple uniform distribution (fixed weighting scheme) produces the most accurate GDP forecasts at all frequencies. To facilitate comparison with the CDFM and the other competing models, the performance of MIDAS regression was measured using the NRMSE and MAE measures of forecast error. Forecasting quarterly GDP using MIDAS regression requires future values of monthly indicators as input. The forecasts of the monthly indicators were determined using optimally selected ARMA models, with model selection based on BIC.

There are several important conceptual differences between a MIDAS regression and a

DFM used as a building block of the CDFM, or as direct competing model in Section C. First, a DFM is based on a system of equations, whereas a MIDAS regression involves a single equation in reduced form. Bai et al. (2013) argue that a MIDAS model can be less efficient and more prone to specification errors than a DFM estimated using the Kalman filter. Second, because a MIDAS regression requires forecasts for the indicators as input, it must be supplemented by auxiliary models for the indicators. This makes the choice of a MIDAS regression less practical than that of a DFM for the purpose of designing a cluster of such models such as the CDFM. Finally, a DFM estimated using the Kalman filter can easily cope with ragged edges and missing observations at different frequencies, as well as changing frequencies within a time series, which poses a problem to MIDAS.

F Additional figures and tables

Table 10: Summary statistics

	min	median	mean	max	sd	skew	ac
Import deflator	-6.00	1.60	1.22	7.67	2.58	-0.18	0.85
Private consumption deflator	-0.24	1.88	1.75	3.37	0.75	-0.64	0.91
Export of goods	-18.32	4.97	5.20	20.72	6.80	-0.83	0.87
Export of services	-10.39	4.14	4.15	19.86	4.67	-0.28	0.62
Manufacturing VA	-15.72	2.89	2.54	13.4	4.49	-1.61	0.80
Investment construction	-9.11	0.44	0.15	7.01	3.28	-0.62	0.74
Construction VA	-11.90	-0.16	-0.06	7.03	3.70	-0.57	0.68
Services VA	-1.96	2.25	2.32	6.90	1.79	-0.10	0.85
Labor income manufacturing	-4.92	2.80	2.74	6.67	2.47	-0.85	0.83
Export deflator	-3.90	1.04	0.93	5.33	1.57	-0.01	0.87
GDP deflator	-0.03	1.67	1.61	2.84	0.64	-0.42	0.77
Capital income	-10.04	3.83	3.70	9.88	3.39	-1.40	0.78
Labor income construction	-2.74	2.39	2.52	9.42	2.66	0.07	0.87
Labor income services	0.00	4.07	3.90	6.27	1.41	-0.58	0.87
Private consumption	-1.68	1.28	1.36	4.04	1.09	-0.03	0.70
Investment equipment	-15.84	2.47	2.12	16.84	5.84	-0.19	0.67
Investment intangibles	-1.46	4.58	5.35	13.89	3.85	0.26	0.83
Tradables VA	-10.03	2.8	2.59	9.94	3.19	-1.50	0.85
Nontradables VA	-2.47	1.05	0.98	3.41	1.10	-0.52	0.76
Exports	-15.46	4.68	4.86	15.95	5.59	-1.08	0.87
Imports	-15.28	4.32	4.02	14.94	4.82	-1.29	0.80
Investment	-19.37	2.34	1.75	15.37	5.03	-1.08	0.57
Labor income	0.33	3.62	3.35	6.12	1.31	-0.17	0.88
GDP	-4.76	2.09	1.84	4.79	1.75	-1.40	0.87

AC stands for the autoregressive coefficient of the first order.

Table 11: Ordering of behavioral models in the CDFM

DFM	Target variable $x_t^{(i)}$	Link variables $\boldsymbol{x}_t^l$	Other variables $\boldsymbol{x}_t$
(1)	Import deflator		Oil price in Euro (Brent) • EU PPI manufacturing
(2)	Consumption deflator	Import deflator (1)	CWI
(3)	Exports of goods		EU PMI • US GDP • EU GDP • Truck mileage
(4)	Exports of services	EU PMI (3) • EU GDP (3) • US GDP (3)	Air passenger volume
(5)	Manufacturing VA	Exports of goods (3) • Truck mileage (3)	DE manufacturing confidence • Manufacturing orders • Industrial production • Manufacturing employment • Manufacturing vacancies
(6)	Investment construction	Truck mileage (3) • Manufacturing VA (5)	Construction expectations
(7)	Construction VA	Truck mileage (3) • Manufacturing VA (5) • Investment construction (6)	Construction employment
(8)	Services VA	Truck mileage (3) • Exports of services (4) • Manufacturing VA (5)	Services employment • Services situation • Services expectations • Services vacancies
(9)	Labor income manufacturing	Manufacturing employment (5)	Manufacturing CWI • Manufacturing foreign orders
(10)	Export deflator	Import deflator (1) • EU GDP (3) • Manufacturing CWI (9) • Manufacturing foreign orders (9)	
(11)	Capital income	Exports of goods (3) • Exports of services (4)	Yield curve (2-10y) • Manufacturing uncertainty
(12)	Labor income construction	Construction expectations (6) • Construction employment (7) • Manufacturing CWI (9)	
(13)	Labor income services	Services employment (8) • Services situation (8) • Manufacturing CWI (9)	
(14)	Private consumption	Labor income manufacturing (9) • Capital income (11) • Labor income services (13)	
(15)	Investment equipment	Manufacturing VA (5) • Manufacturing orders (5) • Manufacturing vacancies (5) • Services VA (8) • Capital income (11)	Manufacturing situation
(16)	Investment intangibles	Exports of goods (3) • Investment equipment (15)	

The behavioral models are estimated sequentially by running the Kalman filter. The link variables are indicated by the number of the DFM model which their forecast is sourced from. *CWI* refers to the collective wage index that results from the wage bargaining process. *Manufacturing uncertainty* refers to a direct measure of uncertainty from the manufacturing sector explained in detail in Glocker and Hölzl (2021). Additional behavioral DFMs are estimated for (i) tradables VA, (ii) nontradables (VA), (iii) tradables deflator and (iv) nontradables deflator.

Table 12: Aggregator models in the CDFM

DFM	Variable y_t	Weighted target variables $x_t^{(i)}$
(17)	Exports	Exports of goods • Exports of services
(18)	Imports	Consumption • Investment construction • Investment equipment • Investment intangibles • Exports
(19)	Investment	Investment construction • Investment equipment • Investment intangibles
(20)	Labor income	Labor income manufacturing • Labor income construction • Labor income services
(21)	Employment	Employment manufacturing • Employment construction • Employment services
(22)	GDP deflator	Consumption deflator • Export deflator • Import deflator • Public sector CWI
(23)	GDP production	Manufacturing VA • Construction VA • Services VA
(24)	GDP expenditure	Consumption • Investment • Exports • Imports
(25)	GDP income	(GDP deflator) • Labor income • Capital income
	GDP TNT	Tradable goods VA • Nontradable goods VA
	GDP deflator (TNT)	Tradables deflator • Nontradables deflator

The GDP deflator is used to express nominal variables (labor and capital income) in real terms. *Public sector CWI* refers to the collective wage index that results from the wage bargaining process in the public sector. Details for the tradable-nontradable goods approach in modeling the production side are provided in Section B of the appendix. The aggregator model for imports is based on the import content of consumption, investment and exports. The import content is calculated using the input-output tables.

Table 13: Factor correlation and MAE by core DFM (2007-2018)

		in-sample (R^2)	out-of-sample (MAE)			
DFM	Variable		3m(1q)	6m(2q)	9m(3q)	12m(4q)
(1)	Import deflator	0.77	0.83	1.66	2.17	2.59
(2)	Private consumption deflator	0.60	0.21	0.34	0.47	0.54
(3)	Export of goods	0.78	1.95	3.31	4.47	5.64
(4)	Export of services	0.57	2.05	2.11	2.12	2.26
(5)	Manufacturing VA	0.93	1.82	2.50	3.23	3.63
(6)	Investment construction	0.52	1.23	1.67	2.00	2.17
(7)	Construction VA	0.73	1.49	2.35	3.00	3.49
(8)	Services VA	0.52	0.57	0.81	0.99	1.19
(9)	Labor income manufacturing	0.83	0.75	1.17	1.61	1.92
(10)	Export deflator	0.97	0.48	0.97	1.31	1.52
(11)	Capital income	0.92	1.67	2.06	2.30	2.61
(12)	Labor income construction	0.75	1.11	1.26	1.40	1.64
(13)	Labor income services	0.87	0.46	0.64	0.86	1.03
(14)	Private consumption	0.56	0.64	0.73	0.8	0.82
(15)	Investment equipment	0.73	3.15	3.44	4.11	4.14
(16)	Investment intangibles	0.98	1.58	3.19	4.80	6.02

MAE refers to the mean absolute error.
The y-o-y growth rates and realizations are expressed in percentage units.

Table 14: MAE by aggregator DFM (2007-2018)

DFM	Variable	3m(1q)	6m(2q)	9m(3q)	12m(4q)
(17)	Exports	1.55	2.51	3.38	4.19
(18)	Imports	1.94	2.7	3.38	4.06
(19)	Investment	3.30	3.68	4.17	4.60
(20)	Labor income	0.36	0.53	0.74	0.96
(21)	Employment	0.28	0.47	0.63	0.77
(22)	GDP deflator	0.29	0.38	0.38	0.37
(23)	GDP production	0.55	0.74	0.94	1.17
(24)	GDP expenditure	0.54	0.83	1.10	1.33
(25)	GDP income	0.62	0.79	0.95	1.17
	GDP average	0.55	0.75	0.96	1.20
			Competing models		
	GDP random walk	0.76	1.14	1.52	1.87
	GDP AR(1)	0.71	1.04	1.25	1.44
	GDP ARMA(2,1)	0.69	1.01	1.15	1.34
	GDP Small DFM	0.64	0.92	1.06	1.32
	GDP Large DFM	0.88	0.98	1.04	1.13
	GDP MIDAS	0.59	0.82	1.05	1.21

MAE refers to the mean absolute error.
The y-o-y growth rates and realizations are expressed in percentage units.

Table 15: Diebold-Mariano test for GDP forecasts (2007-2018)

	Production	Expenditure	Income	TNT	Random walk	AR(1)	ARMA(2,1)	Small DFM	Large DFM
					3m(1q)				
Expenditure	<								
Income	>*	>							
TNT	>	>	<						
Random walk	>***	>**	>*	>					
AR(1)	>***	>*	>	>	<				
ARMA(2,1)	>**	>	>	>	<*	<			
Small DFM	>**	>	<	>	<***	<	<		
Large DFM	>***	>***	>***	>***	>	>	>*	>**	
MIDAS	>	>	<	>	<*	<	<	<	<***
					6m(2q)				
Expenditure	<								
Income	>	>							
TNT	>	>	>						
Random walk	>***	>*	>**	>					
AR(1)	>***	>	>**	>	<				
ARMA(2,1)	>***	>	>**	>	<	<			
Small DFM	>***	>	>*	>	<**	<	<		
Large DFM	>***	>	>***	>	<	<	<	<	
MIDAS	>**	>	>*	>	<	<	<	>	>
					9m(3q)				
Expenditure	>*								
Income	>	>							
TNT	>**	>	>**						
Random walk	>***	>**	>***	>					
AR(1)	>**	>	>**	>	<***				
ARMA(2,1)	>*	>	>*	>	<***	<**			
Small DFM	>	>	>	<	<***	<	<		
Large DFM	<	<	<	<	<**	<	<	<	
MIDAS	>*	>	>	>	<**	<	<	>	>
					12m(4q)				
Expenditure	>**								
Income	<	<*							
TNT	>**	>	>*						
Random walk	>***	>**	>***	>*					
AR(1)	>*	>	>*	>	<***				
ARMA(2,1)	>	>	>	>	<***	<**			
Small DFM	>	>	>	>	<***	<	<		
Large DFM	<	<	<	<	<***	<	<	<	
MIDAS	>	>	>	<	<***	<**	<	<	>

The inequality sign compares a row against a column variable. The loss function is based on absolute deviations.
The notation *** 1 percent; ** 5 percent; * 10 percent level of significance of a two-sided Diebold-Mariano test.

Table 16: Coroneo-Iacone test for GDP forecasts (2007-2018)

	Production	Expenditure	Income	TNT	Random walk	AR(1)	ARMA(2,1)	Small DFM	Large DFM
					3m(1q)				
Expenditure	<								
Income	>	>							
TNT	>	>	<						
Random walk	>**	>**	>**	>					
AR(1)	>**	>	>	>	<				
ARMA(2,1)	>**	>	>	>	<	<			
Small DFM	>**	>	<	>	<**	<	<		
Large DFM	>**	>**	>**	>**	>	>	>**	>**	
MIDAS	>	>	<	>	<**	<	<	<	<**
					6m(2q)				
Expenditure	<								
Income	>	>							
TNT	>	>	>						
Random walk	>**	>	>**	>					
AR(1)	>**	>	>**	>	<				
ARMA(2,1)	>**	>	>**	>	<	<			
Small DFM	>**	>	>	>	<**	<	<		
Large DFM	>**	>	>**	>	<	<	<	<	
MIDAS	>	>	>	>	<**	<**	<**	>	>
					9m(3q)				
Expenditure	>**								
Income	>	>							
TNT	>**	>	>**						
Random walk	>**	>	>**	>					
AR(1)	>**	>	>**	>	<**				
ARMA(2,1)	>	>	>	>	<**	<**			
Small DFM	>	>	>	<	<**	<	<		
Large DFM	<	<	<	<	<	<	<	<	
MIDAS	>	>	>	>	<**	<**	<	>	>
					12m(4q)				
Expenditure	>**								
Income	<	<**							
TNT	>**	>	>						
Random walk	>**	>	>**	>					
AR(1)	>	>	>	>	<**				
ARMA(2,1)	>	>	>	>	<**	<			
Small DFM	>	>	>	>	<**	<	<		
Large DFM	<	<	<	<	<**	<	<	<	
MIDAS	>	>	>	<	<**	<**	<**	<	>

The inequality sign compares a row against a column variable. The loss function is based on absolute deviations.
The notation ** 5 percent level of significance of a Coroneo-Iacone (fixed-b) test.