Forecasting the production side of GDP

Gregor Bäurle, Elizabeth Steiner and Gabriel Züllig

Forecasting the production side of GDP[*]

Gregor Bäurle[†], Elizabeth Steiner[‡] and Gabriel Züllig[§]

December 13, 2018

Abstract

We evaluate the forecasting performance of time series models for the production side of GDP, that is, for the sectoral real value added series summing up to aggregate output. We focus on two strategies that are typically implemented to model a large number of time series simultaneously: a Bayesian vector autoregressive model (BVAR) and a factor model structure; we then compare them to simple benchmarks. We look at point and density forecasts for aggregate GDP, as well as forecasts of the cross-sectional distribution of sectoral real value added growth in the euro area and Switzerland. We find that the factor model structure outperforms the benchmarks in most tests, and in many cases also the BVAR. An analysis of the covariance matrix of the sectoral forecast errors suggests that the superiority of the factor model can be traced back to its ability to capture sectoral comovement more accurately, and the fact that this gain is especially high in periods of large sectoral dispersion.

JEL classification: C11, C32, C38, E32, E37

Keywords: Forecasting, GDP, Sectoral heterogeneity, Bayesian vector autoregression, Dynamic Factor Model

1 Introduction

There is an extensive literature that proposes and evaluates methods for forecasting real GDP. For a long time, researchers concentrated on analysing the most precise forecasts, i.e. 'point' forecasts. More recently, they have turned to a second aspect of forecast analysis, looking at 'density' forecasts, that is estimating the uncertainty contained in GDP forecasts. However, there is also a further aspect, which has been much neglected up to now. Not least the stark, long-lasting policy interventions during and after the financial crisis sparked an interest in the joint evolution of macro-economic variables and sectoral heterogeneity, i.e. in the cross-sectional distribution of the production sectors that together constitute the real economy.

In this paper, we evaluate the forecasting performance of time series models describing value added series for the many production sectors that sum up to aggregate output. A useful production-side model should arguably perform well in all three aspects mentioned above. We therefore evaluate the model forecasting performance comprehensively by assessing the point and density forecasts for aggregate GDP, as well as the cross-sectional distribution of the production sectors. We focus on the forecast performance of different 'macro-econometric', production-side models, i.e. models that are able to capture the joint dynamics of the sectoral series and important macroeconomic variables, relative to several simpler benchmark models.

Our analysis proceeds in three steps. First, we present our set of models and describe how they can be estimated using Bayesian methods. We concentrate on models that are suited for the many sectoral time series jointly with important macroeconomic time series. We therefore include a Bayesian vector autoregressive model (BVAR), which is probably the most popular choice for modelling many macroeconomic time series simultaneously. In addition, taking into account the literature on factor-augmented vector autoregressive models, we propose an alternative approach relying on a dynamic factor model structure (DFM). In short, this approach assumes that each of the sectoral series can be decomposed into a component driven by macro-economic factors and a sector-specific component that is orthogonal to these factors. The macroeconomic factors are modelled as a BVAR, while the sector-specific component follows a univariate process. Our set of models is completed with a number of simpler benchmark models.

In a second step, we evaluate the point and density forecast performance of these models for aggregate GDP using data for the euro area and Switzerland. As sectoral real value added sums up to aggregate GDP, all our sectoral models provide us directly with a prediction for aggregate GDP. The performance evaluation, based on standard evaluation criteria, is straightforward. In addition to the evaluation of standard measures such as the RMSE, we analyse the decomposition of the aggregate forecast error variances into the weighted sum of forecast error variance and the covariances between the sectors. If a model performs better owing to a reduction in the covariance terms, this indicates that the model captures the comovement between the series more accurately. We furthermore analyse the role of sectoral comovement by looking separately at episodes with high and low comovement.

In a third step, we turn to the evaluation of the cross-sectional forecast distribution. For this evaluation, we rely on standard measures for multivariate forecasts such as the log determinant of (weighted) error covariance matrix and a multivariate mean squared error.[1] But we also propose two new measures comparing specific aspects of the forecast distribution. The first measure compares the weighted share of sectors that were correctly projected to grow above and below their long-term average respectively. This criterion reflects the idea that a model is useful if it is able to forecast what stage in the business cycle a sector might be in. The second measure assesses how well models predict the dispersion of growth across the economy as measured by its cross-sectoral standard deviation. Looking at the second moment of the cross-sectional distribution allows us to tell whether a model is able to predict the future dispersion of the sectors, abstracting from its sectoral point forecast performance. In other words, a model can perform well if it correctly predicts how different the sectors are from each other, even if it is not able to tell precisely how each sector will grow.

We find quite distinct evidence that the factor model performs very well, irrespective of the evaluation measure. Indeed, it outperforms the simple benchmarks in most tests, and

---

[1] In the literature, this criterion is also labelled 'weighted trace mean squared forecast error'.

in many cases also the BVAR. This is true for both point and density forecasts. In the latter case, the superiority tends to be even more pronounced. Our decomposition of the forecast error variances into sector-specific variances and covariances between sectors supports the hypothesis that the factor model outperforms its competitors because it is better able to understand the degree of sectoral comovement. Interestingly, this is particularly the case if idiosyncratic factors are important, such that sectoral comovement is low. Moreover, the factor model tends to outperform the other models also at forecasting sectoral heterogeneity. In particular, it more accurately forecasts the sectoral dispersion as measured by the cross-sectional standard deviation of the sectors.

Before turning to the description of the models and their evaluation, we present some remarks on the existing literature. We then show the results of the sectoral heterogeneity analysis.

2 Related literature

Our paper contributes to the strand of literature that compares the forecasting performance of models using aggregate data with those that incorporate disaggregate information. A few papers derive analytical results. A key prediction from the theoretical literature is that an optimal model trades off potential model misspecification in aggregate models and increased estimation uncertainty, due to the higher number of parameters in disaggregate models (see e.g. Hendry and Hubrich (2011)). General theoretical results regarding the determinants of this trade-off are scarce. One exception is an early conjecture by Taylor (1978). He argues, based on analytical considerations, that the trade-off depends on the extent of comovement between the disaggregate series. Models using aggregate series or univariate models for disaggregate series are inefficient if the disaggregate series exhibit heterogenous dynamics. At the same time, gains of multivariate disaggregate models are predicted to be rather small if the series move homogeneously. We assess this hypothesis empirically in our setting in section 5.3.

Given that the relative forecast performance of disaggregate and aggregate models depends on the specifics of the data, a number of papers provide empirical assessments. Marcellino

et al. (2003) propose an indicator model using a geographical disaggregation of GDP for the euro area. Zellner and Tobias (2000) put forward a similar model for a set of 18 countries. Both of these early papers show the superiority of models using disaggregate data. There are, however, only a few studies that look at the production-side disaggregation of GDP. Most of them focus predominantly on the point forecast performance of indicator-based models, and therefore concentrate mostly on short-term forecasts.

For the euro area, Hahn and Skudelny (2008) find that choosing the best-performing bridge equations for each sector of production outperforms an AR model forecasting aggregate GDP directly. Barhoumi et al. (2012) perform a similar analysis for the French economy and reach the same conclusion. Drechsel and Scheufele (2012) analyse the performance of a production-side disaggregation and a disaggregation into the expenditure components of German GDP. They compare the resultant forecasts with those of aggregate benchmarks. Overall, they find only limited evidence that bottom-up approaches lead to better predictions. However, in certain cases the production-side approach produces statistically significantly smaller forecast errors than the direct GDP forecasts. More recently, Martinsen et al. (2014) find that disaggregate survey data at a regional and sectoral level improve the performance of factor models in forecasting overall output growth. Along with these analyses, a vast literature has emerged that tests the optimal number of indicators needed to forecast a specific aggregate target variable. Barhoumi et al. (2012) and Boivin and Ng (2006) provide evidence that a medium-sized number of indicators often leads to better performance than a very large number. This is because idiosyncratic errors are often cross-correlated.

A major caveat of most of these indicator models is, however, that they are not able to capture sectoral linkages and comovement explicitly. Production networks play an important role in the propagation of shocks throughout the economy, and can cause low-level shocks to lead to sizeable aggregate fluctuations, as argued by Horvath (1998) and more recently Carvalho et al. (2016). As sectoral linkages are important amplifiers of aggregate movements, their inclusion in a model should presumably help to improve forecasts of aggregate variables.

A number of studies have measured the forecasting performance of models that take

linkages into account, and have compared these to models with non-disaggregated data. The bulk of them is applied to forecasting inflation, with ambiguous results. Hubrich (2005) simulates out-of-sample forecasts for euro area inflation and its five sub-components, and finds that using disaggregate data does not necessarily help, although there are some improvements on medium-term forecast horizons. The reason is that in the models used, many shocks affect the sub-components of inflation in similar ways. This creates highly correlated errors of the components, which are then added up rather than cancelled out. Additionally, more disaggregation comes at the cost of a higher number of parameters to estimate, with decreasing precision. As a consequence, Hendry and Hubrich (2011) favour forecasting aggregate inflation directly using disaggregate information, rather than combining disaggregate forecasts.

These findings have, however, been refuted by Dées and Günther (2014)'s work. They use a panel of sectoral price data from five geographical areas to forecast different measures of inflation, and find that the disaggregate approach improves forecast performance, especially for medium-term horizons. Bermingham and D'Agostino (2011) emphasise that the benefits of disaggregation increase with the number of disaggregate series, but only when one uses models that pick up common factors and feedback effects, such as factor-augmented or BVAR models.

Based on this literature, we test whether modelling the production side of GDP using models that allow for dynamic linkages is beneficial. To the best of our knowledge, we are the first to do so. We move beyond the evaluation of point forecasts and also test the quality of the density forecasts. The tests are carried out for the short run as well as for the medium run (eight quarters ahead). Furthermore, we assess the accuracy of the sector-level forecasts.

3   Models

For the forecasting of macroeconomic time series, a vector autoregressive model (VAR) is usually a good starting point. Each variable is modelled as a function of its own lags and the lags of all other variables included in the model. Such models can be used for forecasting and also, albeit with some limitations, for more structural analysis. Because we use a large set of variables including macro and sectoral series, some shrinkage of the parameter space is required, as the number of parameters increases quadratically with the number of observed variables in a VAR.

In the literature, there are two popular approaches for achieving a parsimonious, simultaneous modelling of a large number of time series. The first is a BVAR approach, i.e. a Bayesian version of a standard VAR (Litterman, 1979, Doan et al., 1984). The shrinkage of the parameter space is achieved by means of informative priors on the coefficients of the model. The second strategy that has become increasingly popular for modelling a large set of macroeconomic time series and for forecasting is a dynamic factor structure (Stock and Watson, 2002). It is assumed that the comovement between observed series can be described appropriately with few common factors. Each observed series is then a linear combination of these factors and their lags, and an idiosyncratic component. The factors themselves are modelled as a dynamic process,[2] giving it its characteristic name Dynamic Factor Model (DFM).[3]

A strong point of both types of model, the BVAR and the DFM, is that they are able to track down which macroeconomic shocks are driving the economy. Bäurle and Steiner (2015), for example, measure the response of macroeconomic shocks on sector-specific value added within a DFM framework. Such analyses enable us to quantify the impact of aggregate shocks on the individual production sectors of an economy. As the transmission of such shocks often takes a few quarters, in addition to the results for the short run we also analyse the medium-run forecasts (eight quarters ahead). To evaluate how well both

---

[2]The terms 'dynamic' vs 'static' factor models are not used uniformly in the literature. Bai and Ng (2002) refer to a 'dynamic' factor model if the observed series load on the factors and also their lags. Note, however, that such a model can be rewritten in a 'static' form by redefining the state vector.

[3]Note that when the primary interest is not to model the large set of variables per se, but merely to extract information from these variables that can be included in a standard VAR, then the model is usually called a 'factor augmented vector autoregression' (FAVAR, see e.g. Bernanke et al. (2005)).

model types are able to forecast the economy in different dimensions (point and density forecast, as well as sectoral dispersion), we simulate a horse race between them and a set of simpler benchmark models. The latter cannot be used for structural analysis, but are known to perform relatively well for forecasting. Both of the two main models (which operate on the full set of sectoral series) and the simpler benchmark models are described in the rest of this section.

We denote quarter-on-quarter value added growth of a single sector s at time t by x_t^s and the stacked vector of x_t^s in all S sectors by X_t^S. The vector of macro variables is denoted by X_t^M. The vector of X_t^S and X_t^M stacked into one vector is denoted by X_t. This contains all data that is jointly used for the two baseline models. Growth in aggregate GDP y_t equals the weighted sum of sectoral value added growth, $\sum_{s=1}^{S} \omega_{s,t-1} x_t^s$, where the weights $\omega_{s,t-1}$ are the nominal value added share of total value added of the previous time period, as is commonly used to calculate growth contributions.

3.1 Large Vector Autoregressive Model (VAR-L)

We set up a large Vector Autoregressive Model (VAR-L) using all macro variables X_t^M and sectoral value added series X_t^S, and estimate the model with Bayesian methods. The stacked vector X_t is assumed to depend linearly on its lags and some disturbances ε_t:

$$X_t = c + \sum_{k=1}^{p} \Phi_k X_{t-k} + \varepsilon_t \tag{1}$$

where the constant c and $\Phi_k, k = 1, \ldots, p$ are coefficient matrices and ε_t is a vector of innovations, which are assumed to be Gaussian white noise, i.e. $\varepsilon_t \sim N(0, \Sigma)$.

With X_t reaching a large dimension, the number of parameters to be estimated is large, relative to the number of available observations. Thus some shrinking of the parameter space is needed. Following the vast majority of the literature, this is achieved by using a Minnesota type prior. Our implementation follows Banbura et al. (2010) and sets the first and the second prior moments of the elements in the i-th row and the j-th column of Φ_k, $k = 1, .., p$ as follows:

$$E(\phi_{ijk}) = \begin{cases} \delta_i & j = i, k = 1 \\ 0 & j \neq i \end{cases}, \quad V(\phi_{ijk}) = \begin{cases} \frac{\lambda^2}{k^2} & j = i \\ \vartheta \frac{\lambda^2 \sigma_i^2}{k^2 \sigma_j^2} & j \neq i \end{cases}$$

The prior distribution implements the uncertain belief that the first own lag of each series i is δ_i and the other coefficients are zero, where the uncertainty with respect to cross-variable coefficients (i.e. the coefficient relating the series i to a lag of series j, $i \neq j$) is proportional to the relative variance of the residuals for the respective variables. The tightness of the 'own' coefficients relative to the 'cross-variable' coefficients is scaled with an additional factor ϑ. Importantly, the uncertainty decreases with the lag length k, making feasible the specification of models with large lag length. The overall tightness of the prior is controlled by the scale parameter λ.

3.2 Dynamic factor Model (DFM)

The DFM framework relates a large panel of economic indicators to a limited number of observed and unobserved common factors. The premise behind this type of model is that the economy can be characterised by a small number of factors that drive the comovements of the indicators in the panel. Rather than summarising indicator data by means of factor analysis, we use it to extract information contained in sectoral value added series by including them in the dynamic system. Formally, the model consists of two different equations: an observation equation and a state equation. The observation equation relates sectoral value added growth X_t^S to the common factors f_t that drive the economy:

$$X_t^S = c + \sum_{k=1}^{p} \Lambda_k f_{t-k} + u_t, \tag{2}$$

where $\Lambda_k, k = 1, \ldots, p$ are the factor loadings and u_t is a vector of item-specific components. Thus, X_t^S is allowed to load on the factors both contemporaneously and on their lags. Importantly, f_t consists of both unobserved and observed factors.[4] In our case, the observed factors are the macro variables X_t^M. Following Boivin and Giannoni (2006), we

---

[4]In order to estimate the model, we rewrite the model in a static state space form. Observed factors are treated as 'unobserved factors' without noise in the observation equation.

allow u_t to be autocorrelated of order one by specifying $u_t = \psi u_{t-1} + \xi_t$ with $\xi_t \sim N(0, R)$. The joint dynamics of the factor f_t are described by the following state equation:

$$f_t = \sum_{k=1}^{p} \Phi_k f_{t-k} + \varepsilon_t \tag{3}$$

where $\Phi_k, k = 1, \ldots, p$ are coefficient matrices, ε_t is a vector of white noise innovations, i.e. $\varepsilon_t \sim N(0, \Sigma)$. Moreover, ε_t and the idiosyncratic shocks u_t are assumed to be uncorrelated.

The model is estimated using Bayesian methods. Since it is not possible to derive analytical results for high-dimensional estimation problems such as the one at hand, we have to rely on numerical techniques to approximate the posterior. In particular, we use a Gibbs Sampler, iterating over conditional draws of the factors and parameters. A detailed account of the step-by-step estimation algorithm is provided in appendix 9.1.

Our choices for the prior distributions are the following. The prior for the coefficients in the observation equation, Λ_k, is proper. This mitigates the problem that the likelihood is invariant to an invertible rotation of the factors. The problem of rotational indeterminacy in this Bayesian context is discussed in detail in Bäurle (2013).[5] We assume that, a priori, the variances of the parameters in Λ_k are decreasing with the squared lag number k, in analogy to the idea implemented in the Minnesota prior that longer lags are less important. The determination of the coefficients describing the factor dynamics reduces to the estimation of a standard VAR. We assume a Minnesota-type prior for the parameters in the state equation.

3.3 Benchmark models

Four different benchmarks, two sectoral and two aggregate ones, complete our suite of models used in the horse race. All of them can be formulated as a special case of the VAR described in Section 3.1. The Bayesian estimation procedure can thus be directly applied, and we get a distribution of forecasts for each of the models. This enables us to evaluate the density forecasts.

The first benchmark model is a combination of VARs, which include the baseline macro variables, X_t^M, plus one sectoral value added series x_t^s and the aggregate of the remaining sectors X_t^{-s}. We first estimate this model for each sector separately and then aggregate the sector forecasts to compute GDP, using nominal value added weights of the last available time period, $\omega_{s,t-1}$.

$$\begin{pmatrix} x_t^s \\ X_t^{-s} \\ X_t^M \end{pmatrix} = c + \sum_{k=1}^{p} \Phi_k \begin{pmatrix} x_{t-k}^s \\ X_{t-k}^{-s} \\ X_{t-k}^M \end{pmatrix} + \varepsilon_t \tag{4}$$

This model has been used e.g. by Fares and Srour (2001) for Canada and Ganley and Salmon (1997) for the UK to analyse the impact of monetary policy at the sectoral level. We label it VAR-S to highlight the sectoral component, as it takes into account the heterogeneity of sectors responding to macroeconomic conditions and shocks.[6]

The second benchmark model, called VAR-A, is the direct aggregate counterpart to VAR-S. This model differs only with respect to the target variable such that it includes GDP y_t directly as a variable in the dynamic system.

$$\begin{pmatrix} y_t \\ X_t^M \end{pmatrix} = c + \sum_{k=1}^{p} \Phi_k \begin{pmatrix} y_{t-k} \\ X_{t-k}^M \end{pmatrix} + \varepsilon_t \tag{5}$$

Ultimately, we have included two univariate AR processes: The AR-S estimates an independent sectoral process and makes predictions which are then aggregated up, equivalent to VAR-S.

$$x_t^s = c + \sum_{k=1}^{p} \phi_{k,s} \, x_{t-k}^s + \varepsilon_t \tag{6}$$

The AR-A is again the aggregate counterpart, which has a minimal number of parameters

---

[5]Bayesian analysis is always possible in the context of non-identified models, as long as a proper prior on all coefficients is specified, see e.g. Poirier (1998). Note that rotating the factors does not impact the impulse response functions as long as no restrictions are set on the responses of the factors to shocks.

[6]This model shows similarities to a "Global VAR" as proposed by Pesaran and Weiner (2010). It actually corresponds to a Global VAR in which the weight in the aggregation is the sectoral share in aggregate GDP, as opposed to weights based on patterns of trade as is typical in Global VARs that model different countries or regions. An alternative weighting scheme in the case of sectoral variables could be based on input-output tables. Due to data limitations, we do not pursue this avenue.

and is a natural choice as a simple but competitive benchmark for forecasting GDP.

$$y_t = c + \sum_{k=1}^{p} \varphi_k \, y_{t-k} + \varepsilon_t \qquad (7)$$

3.4 Specification

We set the number of lags to four for all models. The relative point forecast performance neither increases nor deteriorates systematically when using only one lag instead, but density forecasts tend to worsen.

The prior means, δ_i, are set to zero in the specification for the autoregressive coefficients. Following Banbura et al. (2010), the factor ϑ, controlling the relative importance of other lags relative to own lags, is set to one. This allows us to implement the Minnesota prior with a (conjugate) normal inverted Wishart prior (see e.g. Karlsson (2013)). The overall scaling factor of the prior variance, λ, is chosen according to recommendations by Banbura et al. (2010) based on an optimisation criterion for VARs of similar size, and summarised in Table 1.[7] We take 20,000 draws from the posterior distribution, whereas in the DFM case, we discard an additional 2,000 initial draws to alleviate the effect of the initial values in the MCMC algorithm.

We assess properties of the estimation algorithm in the DFM case using a set of different diagnostics. Geweke's spectral-based measure of relative numerical efficiency (RNE, see e.g. Geweke (2005)) suggests that efficiency loss of the algorithm due to the remaining autocorrelation in these evaluated draws is minimal.[8] The efficiency loss is less than 50 percent for almost all the parameters, i.e. vis-à-vis a hypothetical independence chain, we need no more than half the amount of additional draws to achieve the same numerical precision. Moreover, with a value between 7 and 9, depending on the sample, the maximum inverse RNE is well below the critical threshold discussed in the literature (see e.g. Carriero et al. 2014, Baumeister and Benati 2013 or Primiceri 2005). Additionally, we investigate

---

[7]Note that in principle, it is possible to estimate the weight based on marginal data densities (Giannone and Primiceri, 2015). As we re-estimate our models many times in our forecasting evaluation, and the calculation of the marginal data density is not available in an analytical form in the DFM case, we refrain from this. A numerical approximation to the marginal data density is possible in principle, but the accuracy of such estimators deteriorates with growing dimensionality of the parameter space. See e.g. Fuentes-Albero and Melosi (2013) for a Monte Carlo study and Bäurle (2013) for an application.

[8]The spectrum at frequency zero is calculated using a quadratic spectral kernel (Neusser, 2009).

convergence visually by looking at the posterior means based on an expanding number of draws, finding no evidence of changes after less than half of the draws.

Table 1 summarises the models and their specifications.

TABLE 1: EVALUATED MODELS AND THEIR SPECIFICATIONS: OVERVIEW

	DFM	VAR-L	VAR-S
Description	Dynamic factor model w/ all sectoral and macro series	Large BVAR w/ all sectoral and macro series	VAR w/ one sectoral series at a time and macro series
Real variables	SVA	SVA	SVA
# estimated models	1	1	S
# total variables	S+M	S+M	1+M
Forecast	WS	WS	WS
λ	0.2	0.1	0.2
Reference equation	(2),(3)	(1)	(4)

	VAR-A	AR-S	AR-A
Description	Aggregate VAR w/ GDP and macro series	Univariate AR w/ one sectoral series at a time	Univariate AR w/ GDP
Real variables	GDP	SVA	GDP
# estimated models	1	S	1
# total variables	1+M	1	1
Forecast	D	WS	D
λ	0.2	large	large
Reference equation	(5)	(6)	(7)

Note: (SVA) Sector value added series, (S) The number of sectors, (M) The amount of macro variables. *Forecast* describes how the aggregation of the forecasts to GDP is done: (WS) GDP is the weighted sum of the sectoral growth forecasts, (D) GDP growth is directly forecast.

4 Data

We fit the models to production-side national account data for Switzerland and the euro area. Real value added time series on a quarterly frequency are provided by the Swiss Confederation's State Secretariat for Economic Affairs (SECO, starting in 1990) and Eurostat (starting in 1996) respectively. In contrast to the quarterly GDP series for

the US, the estimation of GDP in Switzerland and in the euro area are both calculated as the sum of the individual production sectors. Switzerland publishes the production-side at a slightly more disaggregate level than Eurostat. For instance, banking and insurance services are reported as separate accounts in Switzerland (and together account for a tenth of GDP) but are merged together in the euro area (where the equivalent share is less than 5 percent of GDP). Overall, the models include a diversified set of industry and services sectors - 13 for the Swiss models and 10 for the European models - which together sum up to GDP.[9] A full descriptive summary of the sectoral series, their volatility, correlation with GDP and autocorrelation is documented in Tables 2 and 3.

These two tables also report the descriptive statistics for total GDP growth rates. The aggregate picture is very similar for Switzerland and the euro area: In the estimation sample, the mean of quarterly GDP growth was 0.38 and 0.39 percent, respectively. Euro area GDP growth is only marginally more volatile. The persistence of aggregate growth rates, measured as the first-degree autocorrelation, is higher in the euro area, but overall, the aggregate characteristics of both GDP series display a high degree of similarity. Figure 1 shows, however, that the downturn during the Great Recession was much more severe in the euro area than in Switzerland.

At the disaggregated sectoral level, the Swiss series are, without any exception, more volatile than their euro area counterparts and less correlated with the aggregate dynamics, indicating that sector-specific features play a larger role in Switzerland. Manufacturing production, typically a sector that shows a high correlation with GDP, is the only series with a contemporaneous correlation coefficient higher than 0.50.

Besides the growth path of aggregate GDP, Figure 1 shows the time series of cross-sectoral standard deviations of sectoral value added growth. Cross-sectoral dispersion is higher in Switzerland throughout the entire estimation sample. It tends to be countercyclical; dispersion typically peaks in recession episodes. Furthermore, the strongest and weakest growing sectors for every quarter are displayed in grey. While these tail sectors are closely related to aggregate dynamics in the euro area, the divergence in Switzerland is striking.

FIGURE 1: GDP GROWTH AND ITS DISPERSION: TIME SERIES
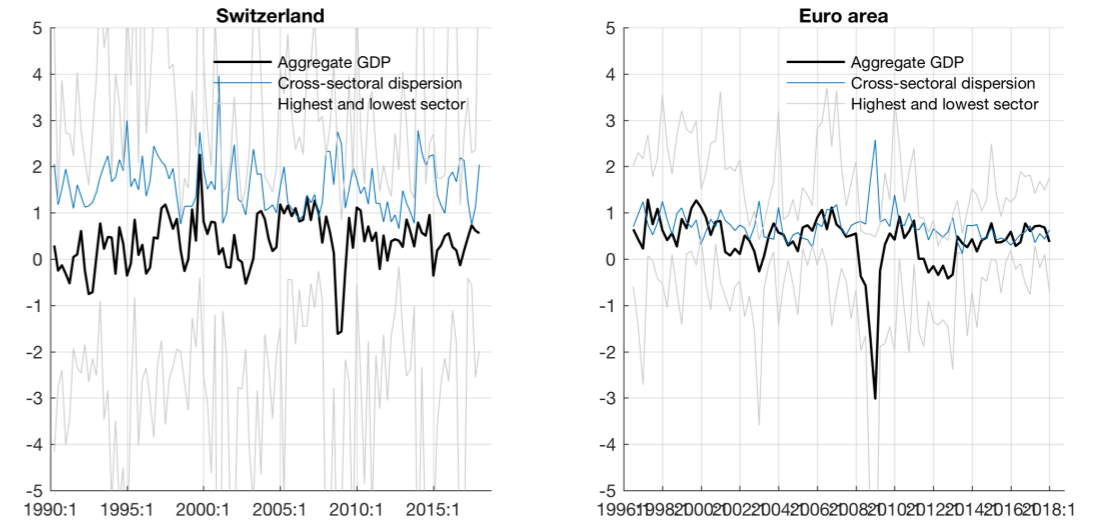
Note: Aggregate vs. disaggregate time series in Switzerland and the euro area. The cross-sectoral standard deviation in a given quarter is displayed in blue; both the highest and lowest sectoral growth rates in the respective time period are given as grey.

Another measure of comovement can be obtained by computing each sector's correlation with aggregate value added, as in Christiano and Fitzgerald (1998), weighted by the respective nominal shares. We repeat the computation in a rolling window of two years and get a time-varying estimate of sectoral comovement that is displayed in Figure 2.[10]

While the level of comovement is on average higher in the euro area, it is also subject to stronger fluctuations. In both economic areas, regimes of high and low comovement can be identified. In general, recessions are associated with high comovement, indicating that economic contractions often affect a large share of the production sectors, but booms can as well, as the economy expands on a broad base. These series allow us to assess whether the forecasting performance varies with the degree of comovement in the target variables.

In addition to the value added series, a set of observable macro factors enters the system of equations. Key economic variables include inflation (CPI and HICP, in log differences) and the nominal short-term interest rates (CHF Libor and Euribor). As Switzerland is a small open economy, in the Swiss models we add a nominal effective exchange rate vis-à-vis its most important trading partners as well as a measure of world GDP. Both series are

---

[9]An advantage of using the production side to forecast GDP is that it is not necessary to produce a forecast for the inventories, which are often not explicable and therefore hardly predictable.

[10]The comovement time series is defined as $\rho_t^c = \sum_{s=1}^{S} \omega_{s,t-1} \rho_t^s$, where ρ_t^s is the correlation coefficient of GDP and the growth rates of each sector s over the 8 quarters subsequent to time period t.
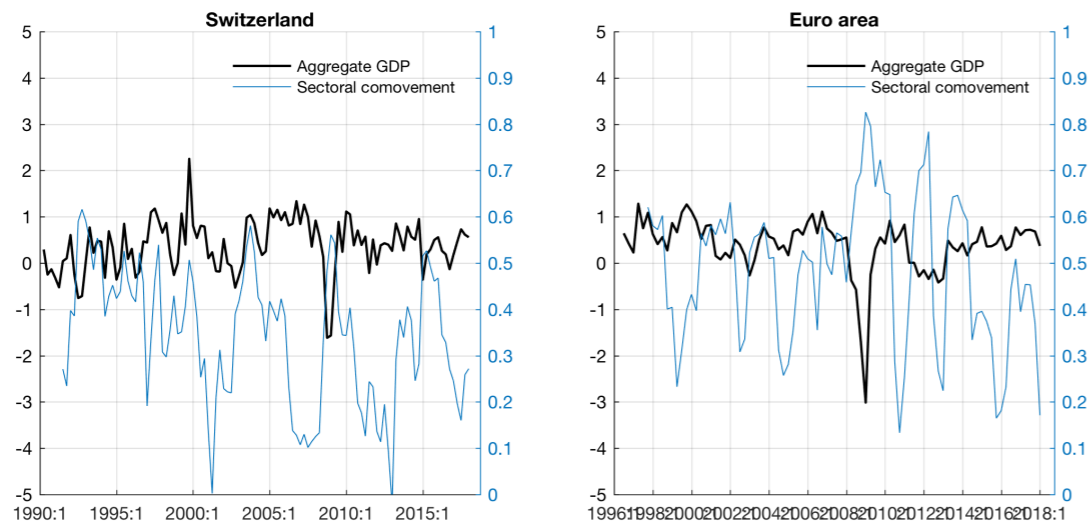
Note: Aggregate vs. disaggregate time series in Switzerland and the euro area. Sectoral comovement is calculated as a weighted correlation coefficient of each sectoral value added with the aggregate, with a rolling window of 8 quarters.

weighted with respect to exports and are defined in log differences.

Our evaluation is based on pseudo out-of-sample forecasts because the availability of real-time vintages is too limited. We use a dataset based on the first quarterly vintage of 2018, which contains data between 1990-Q1 and 2017-Q4 for Switzerland and 1996-Q1 and 2017-Q4 for the euro area. The next sections describe the evaluation exercise and the results in detail.

Table 2: Sectoral value added growth Statistics for Switzerland

Variable	Share	Mean	Std	Corr	Auto
GDP	100.00	0.38	0.57	1.00	0.50
Manufacturing (10-33)	19.44	0.48	1.66	0.73	0.23
Energy (35-39)	2.36	-0.08	2.72	0.11	0.10
Construction (41-43)	5.50	-0.01	1.42	0.14	0.21
Trade, repair (45-47)	14.16	0.47	1.07	0.49	0.66
Transportation, ICT (45-53, 58-63)	8.38	0.33	1.13	0.38	0.60
Tourism, gastronomy (55-56)	2.02	-0.08	2.06	0.33	0.31
Finance (64)	6.07	0.51	3.52	0.43	0.34
Insurance (65)	4.33	1.02	0.87	0.13	0.83
Professional services (68-75, 77-82)	15.31	0.29	0.57	0.36	0.56
Public administration (84)	10.39	0.26	0.42	0.19	-0.00
Health, social services (86-88)	6.46	0.70	0.83	0.16	0.20
Recreation, other (90-96)	2.10	-0.07	2.38	0.14	0.45
Taxes (+) and subsidies (-)	3.47	0.45	0.93	0.39	0.36

Note: NOGA codes in brackets. Share is the average of nominal sectoral value added as a share of GDP between 1990 and 2017. Mean and standard deviation of quarterly log differences, as well as their correlation with aggregate real GDP growth and first-degree autocorrelation.

Table 3: Sectoral value added growth statistics for the euro area

Variable	Share	Mean	Std	Corr	Auto
GDP	100.00	0.39	0.59	1.00	0.66
Industry (C-E)	19.01	0.40	1.47	0.88	0.54
Construction (F)	5.16	-0.09	1.09	0.59	0.29
Trade, transport, tourism (G-I)	17.46	0.42	0.76	0.92	0.54
ICT (J)	4.22	1.21	1.11	0.54	0.35
Finance, insurance (K)	4.52	0.32	1.01	0.34	0.28
Real estate (L)	9.82	0.41	0.45	0.24	-0.03
Professional services (M-N)	9.37	0.54	0.96	0.82	0.40
Public administration (O-Q)	16.98	0.28	0.18	0.26	0.21
Recreation, other (R-U)	3.17	0.29	0.49	0.57	0.40
Taxes (+) and subsidies (-)	10.27	0.31	0.94	0.63	0.07

Note: NACE Rev.2 codes in brackets. Share is the average of nominal sectoral value added as a share of GDP between 1996 and 2017. Mean and standard deviation of quarterly log differences, as well as their correlation with aggregate real GDP growth and first-degree autocorrelation.

5 Evaluation of point forecasting performance

We conduct an out-of-sample forecast evaluation exercise where we assess the models' accuracy in terms of predicting growth in the aggregate.

Out-of-sample forecasts are produced for the twelve years between 2005-Q1 and 2016-Q4. The size of the training data sample (1990-Q1 to 2005-Q1 for Switzerland and 1996-Q1 and 2005-Q1 for the euro area) is sufficient to produce stable estimation results. The last year of the sample is cut off, because end-of-sample data is often subject to substantial future revisions and should not be interpreted as the final vintage (Bernhard, 2016). For this reason, 48 complete vintages are evaluated.

As the models are geared toward capturing complex, dynamic interlinkages in the national accounts, we do not focus on the predicted growth in any specific quarter h periods in the future, but want to assess the models' capability to forecast *cumulative* growth over a range of quarters. For the short run, we produce iterated forecasts for the first four periods ahead; the cumulative sum commensurates to a year-on-year growth rate. The respective evaluation for the second year to be forecasted, which consists of the projected growth from 5 to 8 quarters ahead, is denoted the medium run. This resembles the forecasts conducted by the ECB Survey of Professional Forecasters (SPF), where survey respondents are asked to provide forecasts over a *rolling horizon*, that is an annual growth rate for the quarter one (two) years ahead of the latest available observation.

If y_t is the log difference of the realised target variable from $t = 1, ..., T$, then the cumulative growth over four quarterly periods is denoted as $\tilde{y}_{t|t-h} = \sum_{i=h-4}^{h-1} y_{t-i}$. Accordingly, $\hat{y}_{t|t-h}$ is the cumulative sum of the four steps leading up to the h quarters ahead. Errors are then defined as the difference from the cumulated quarterly growth rates, $e_{h,t} \equiv \hat{y}_{t|t-h} - \tilde{y}_{t,t-h}$.

For the euro area, forecasts from the Survey of Professional Forecasters (SPF), conducted and published by the European Central Bank, are used as an additional benchmark. Note that all models under evaluation fight an uphill battle against the Survey of Professional Forecasters due to the frequency of real-time data releases. Given that all models described rely on national accounts data only, forecasts for all quarters ahead can be updated approximately 30 days into the quarter, when the first estimate is released. The SPF benchmark, however, is produced almost a quarter later. As the SPF not only relies on national accounts data but also on respondents' judgement of a set of early indicators, this extra quarter gives the forecasters a sizeable informational advantage. As an illustrative example, a sample of around 50 respondents to the survey submit their forecast in early 2015-Q1, at which point national accounts data for the preceding quarter have not yet been published, making 2015-Q3 the target period. The rolling forecast horizon for *one year ahead of the latest available observation* therefore effectively implies a forecast horizon of only 3 quarters, giving the SPF a considerable head start.

The following section presents and discusses the relative performance of the competing models. The absolute performance, where every model's capabilities in terms of bias and efficiency of short-run forecasts are evaluated by means of Mincer-Zarnowitz regressions, is displayed in the appendix.

5.1 Relative performance of aggregate forecasts

The relevant metric by which we compare errors across models is the square root of the mean squared error (RMSE):

$$RMSE_{m,h} = \sqrt{\frac{1}{T} \sum_{t}^{T} e_{m,h,t}^2}$$

The respective measures are displayed as part of Table 4 and summarised in Figure 3. To keep the representation of the results tractable, we do not report measures for all forecast horizons separately, but restrict ourselves to two horizons: The short run (SR) is the cumulative forecast error over the first two quarters and the medium run (MR) is the cumulative forecast error over eight quarters. As the uncertainty to forecast a path over several quarters increases in h, so does the expected forecast error.

Furthermore, we use a test following Diebold and Mariano (1995) to assess whether the difference of squared errors of a given model and that of a simple benchmark is statistically significant. As a benchmark we use the simplest of our models, the autoregressive process

Figure 3: Root mean squared errors compared
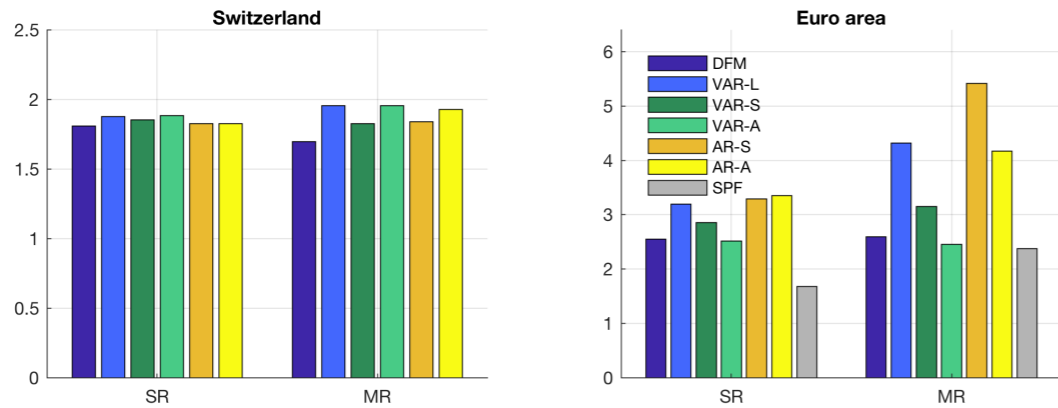
of the aggregate target variable, AR-A.

$$e^2_{m,h,t} - e^2_{AR\text{-}A,h,t} = \beta_{DM} + u_t, \qquad H_0 : \beta_{DM} = 0$$

Table 4 contains the results. If the regression coefficient is negative, the respective model has beaten the benchmark on average over the evaluation period.

Table 4: RMSE and Diebold-Mariano test coefficients

		DFM	VAR-L	VAR-S	VAR-A	AR-S	AR-A
Switzerland							
RMSE	SR	1.81	1.88	1.85	1.89	1.83	1.83
	MR	1.70	1.96	1.83	1.96	1.84	1.93
Diebold-Mariano: β_{DM}	SR	-0.06	0.19	0.10	0.21	-0.00	-
		(0.46)	(0.76)	(0.86)	(0.83)	(0.22)	-
	MR	-0.84	0.11	-0.38	0.10	-0.33	-
		(0.43)	(0.46)	(0.44)	(0.69)	(0.10)	-
Euro area							
RMSE	SR	2.55	3.19	2.85	2.52	3.29	3.36
	MR	2.60	4.32	3.15	2.47	5.39	4.18
Diebold-Mariano: β_{DM}	SR	-4.77	-1.08	-3.12	-4.91	-0.43	-
		(3.90)	(1.66)	(1.21)	(2.61)	(0.57)	-
	MR	-10.75	1.16	-7.49	-11.40	11.58	-
		(11.66)	(2.32)	(6.85)	(10.26)	(10.69)	-

Note: Newey-West standard errors in brackets

For Switzerland, the different models produce short-run forecasts that are not significantly different from each other.[11] With a longer forecast horizon, a significant pattern emerges: The DFM has the lowest errors over the medium run, with an improvement of 12 percent relative to the aggregate AR. According to our test procedure, this difference is significant at the 10 percent significance level. In contrast, the VAR-L, which relies on the same variables but does not impose the factor structure, performs substantially worse than the DFM. This indicates that shrinking the parameter space by using the factor model proves to be crucial for medium-run forecast performance.

Among the simpler benchmarks, the RMSE show that in Switzerland, where sectoral comovement is relatively weak, using disaggregated series helps to improve the medium-run forecast: Both the sectoral AR-S and VAR-S beat their aggregate counterparts.

Errors for euro area GDP forecasts are generally higher, especially in the medium run. This can partly be explained by a limited training sample for estimation and the fact that the downturn during the Great Recession was much more severe in the euro area than in Switzerland and that such strong fluctuations are difficult for any model to capture. This is especially true for the univariate models. Indeed, both in the short run as well as in the medium run, the AR-A and AR-S models perform badly for the euro area.

In the presence of such a large economic shock, more sophisticated models provide superior results. The short-run forecasts of the DFM and of both VAR benchmarks have errors that are 24 percent lower than the aggregate AR. All models perform substantially worse than the SPF in the short run. This is not surprising given that, as mentioned above, it is produced almost one quarter later, when it is possible to exploit evidence from a broader set of (monthly) economic indicators that correlate with GDP growth. In the medium run, the DFM and the VAR-A are competitive with the SPF. These models perform better than simple benchmarks, even if it is difficult to establish an improvement in terms of statistical significance.

Overall, these findings show that including sector information can lead to more accurate point estimates. The best performance, however, comes from the DFM model, which

---

[11]The RMSE for $h = 1$ are depicted in the appendix for reference. The results show that for one quarter ahead, the DFM and the VAR-S produces the best results.

simultaneously models the sectoral value added series and macroeconomic variables while shrinking the parameter space by imposing a factor structure.

5.2 Decomposition of the forecast error variance

One strength of the multivariate models is that they are, in principle, able to capture joint dynamics between sectors. In the below section, we investigate whether the differences in performance documented previously are indeed driven by differential capabilities to capture the joint dynamics. In order to do so, we exploit the fact that in the case of the sectoral models, the aggregate error is a weighted sum of the sectoral errors, and decompose its variance into the sum of the sectoral forecast error variances and covariances:[12]

$$
\mathrm{Var}(e^y) = \mathrm{Var}\left(\sum_{s=1}^{S} \omega_s e_s\right)
$$
$$
= \sum_{s=1}^{S} \omega_s^2 \mathrm{Var}(e_s) + 2 \sum_{1 \le s < \varsigma \le S} \omega_s \omega_\varsigma \mathrm{Cov}(e^s, e^\varsigma) \tag{8}
$$

Table 8 and Figure 4 document the two components of the variance of the aggregate error in equation (8): the weighted sum of the sectoral errors and the covariances between errors of different sectors. The decomposition for the aggregate models is computed by replacing the sectoral prediction with the aggregate prediction. That is, we assume that all sectoral forecasts grow at the same rate, equal to the aggregate one. In Figure 4, the contribution of the sectoral errors is depicted in grey bars, while the sum of the covariances is shown in the colour of the respective model.

In both economic areas, the contribution of the covariance of sectoral errors decreases in absolute and relative terms when sectoral information is included in the models. For the medium-term forecasts in Switzerland, the error due to the covariance term of the decomposition is reduced by a third (from 0.15 to 0.10). In contrast, the sectoral error variance does not vary much across the different models. The VAR-L is almost equally as successful in capturing sectoral covariance, despite having larger aggregate errors. Using

---

[12]Note that the mean squared GDP forecast error in the previous section is the sum of the variance of the forecast error and the squared bias. Mincer-Zarnowitz regressions suggests that the bias is small, see 9 in the appendix.

FIGURE 4: DECOMPOSITION OF ERROR VARIANCES INTO A SECTORAL ERROR COMPONENT (GREY) AND A COMOVEMENT COMPONENT (COLOURISED)
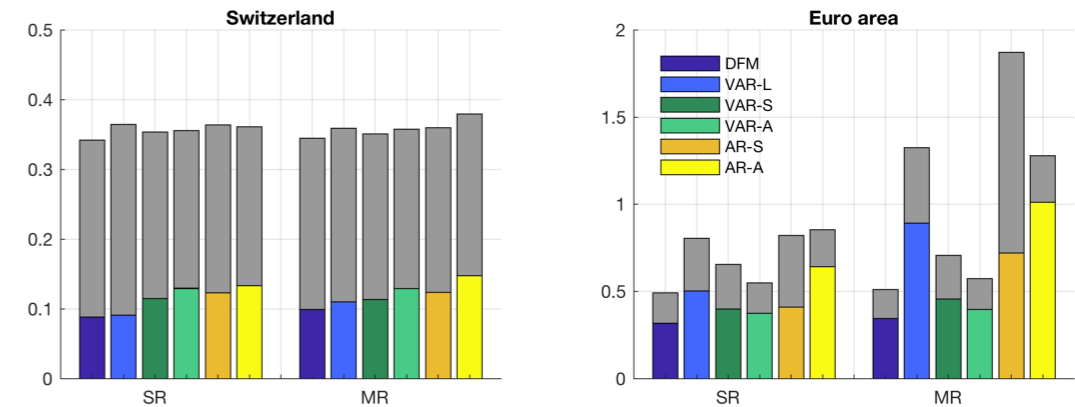
euro area data, the reduction in variance of the aggregate error due to lower covariance of sectoral errors is more distinct. The sectoral models produce substantially lower covariance terms. The differences between aggregate and disaggregate benchmark models can be attributed to the differences in the information set and also to the quite rudimentary construction of the sector forecasts within the aggregate model.

The DFM produces the lowest sectoral error covariance in both economic areas. The figures in Table 8 show that the superior performance of the DFM is mainly due to a reduction in the covariance term, and not primarily to better forecasts for single sectors. This suggests that by modelling sectoral comovement using a factor structure improves the forecast error by reducing the forecast error covariance.

5.3 The role of sectoral comovement

This section assesses the influence of sectoral comovement on the accuracy of point forecasts. Taylor (1978) argues, based on analytical considerations, that models using aggregate series or univariate models for disaggregate series are inefficient if the disaggregate series exhibit heterogeneous dynamics. At the same time, gains of multivariate disaggregate models are predicted to be rather small if the series move homogeneously. Therefore, we would expect that the differences between our models are especially pronounced in periods

TABLE 5: DECOMPOSITION OF THE FORECAST ERROR VARIANCE

		DFM	VAR-L	VAR-S	VAR-A	AR-S	AR-A
Switzerland							
Variance of aggregate errors	SR	0.34	0.36	0.35	0.35	0.36	0.36
Sectoral error variance	SR	0.25	0.27	0.24	0.23	0.24	0.23
as a share		(0.74)	(0.75)	(0.67)	(0.64)	(0.66)	(0.63)
Sectoral error covariance	SR	0.09	0.09	0.12	0.13	0.12	0.13
as a share		(0.26)	(0.25)	(0.33)	(0.36)	(0.34)	(0.37)
Variance of aggregate errors	MR	0.34	0.36	0.35	0.36	0.36	0.38
Sectoral error variance	MR	0.25	0.25	0.24	0.23	0.24	0.23
as a share		(0.71)	(0.69)	(0.68)	(0.64)	(0.66)	(0.61)
Sectoral error covariance	MR	0.10	0.11	0.11	0.13	0.12	0.15
as a share		(0.29)	(0.31)	(0.32)	(0.36)	(0.34)	(0.39)
Euro area							
Variance of aggregate errors	SR	0.49	0.81	0.66	0.55	0.82	0.85
Sectoral error variance	SR	0.17	0.30	0.26	0.17	0.41	0.21
as a share		(0.35)	(0.38)	(0.39)	(0.31)	(0.50)	(0.25)
Sectoral error covariance	SR	0.32	0.50	0.40	0.38	0.41	0.64
as a share		(0.65)	(0.62)	(0.61)	(0.69)	(0.50)	(0.75)
Variance of aggregate errors	MR	0.51	1.33	0.71	0.58	1.86	1.29
Sectoral error variance	MR	0.17	0.43	0.25	0.12	1.14	0.27
as a share		(0.32)	(0.33)	(0.36)	(0.20)	(0.61)	(0.21)
Sectoral error covariance	MR	0.35	0.89	0.46	0.40	0.72	1.02
as a share		(0.68)	(0.67)	(0.64)	(0.69)	(0.39)	(0.79)

of low comovement. In order to assess this hypothesis, we divided the evaluation period into periods of high and low comovement.[13]. We calculate the RMSE on the subsample of errors from high and low comovement periods respectively.

TABLE 6: RMSE IN HIGH/LOW SECTORAL COMOVEMENT REGIMES

		DFM	VAR-L	VAR-S	VAR-A	AR-S	AR-A
Switzerland							
Low comovement	MR	1.31	1.95	1.79	2.01	1.60	1.73
High comovement	MR	2.01	1.97	1.86	1.90	2.05	2.11
Euro area							
Low comovement	MR	1.42	5.32	3.27	1.86	6.89	4.98
High comovement	MR	3.39	3.00	3.02	2.93	3.35	3.17

Table 6 shows the results in high and low comovement regimes for the medium term. Systematic differences in the short-term forecasts could not be found. The results in

---

[13]This allocation into a 'high' or 'low' period is defined using the comovement measure as in Christiano and Fitzgerald (1998) (see Figure 2)

Table 6 show that the relative model ranking presented in section 5.1 is indeed driven by low comovement periods to a large extent. The differences between the models is much less distinct in times of high comovement.

In low comovement regimes, estimating models at the sectoral improves the medium-term forecasts. As with the results in section 5.1, the univariate AR models in the euro area are an exception. Here, the aggregate process performs better than the sectoral one. Furthermore, the VAR-L performs poorly, while the DFM, which not only includes the sectoral series jointly but also manages to filter relevant information at the disaggregate level, produces the most accurate forecasts.

In contrast, in times of high comovement, i.e. when the sectoral idiosyncratic factors are less important and the sectors develop more homogeneously, the gain of disaggregation is much smaller. The sectoral approach does not lead to a systematic improvement in the RMSE. Interestingly, the relative forecasting performance of the VAR-L improves markedly in times of high comovement, while the DFM loses its comparative advantage. By contrast, we find no systematic differences between times of high vs low growth rates or high vs low volatility of quarterly GDP growth rates.[14]

All in all, this exercise shows that when the heterogeneity across sectors is high, [15] models including sectoral series perform better than their aggregated counterparts. During such periods, the DFM produces the most accurate forecasts, in line with the original hypothesis of Taylor (1978).

6 Evaluation of density forecasting performance

Point forecasts do not capture the uncertainty around which a central prediction is made. Density forecasts have, therefore, become an increasingly popular tool to communicate how likely it is that the predictions will fit the realisation. We devote this section to the

---

[14]An analysis of the time variation in the forecasting performance along the lines of Rossi (2013) turned out not to deliver major insights due to the short sample; see Figure 7 in the Appendix. However, cumulated errors (Figure 8) reveal that the DFM's good performance is attributable to the period before the financial crisis, while it declines somewhat in the aftermath.

[15]Regimes of high heterogeneity can, for example, include periods around turning points at the peaks of a business cycle, as documented by (Chang and Hwang, 2015) for the US manufacturing industries.

evaluation of the predictive densities of our models. For each model m, we simulate from the Bayesian posterior distribution of the forecasts in order to determine the density of the cumulative forecast $\phi(\hat{y}_{m,t|t-h})$.

The fundamental problem of evaluating density forecasts in contrast to point forecasts is that the actual density is unobserved, i.e. we observe just one realisation, not many realisations of the same density. A number of methods have been developed to address this. These include the probability integral transforms (PIT), evaluations based on the log score and, related to this, the ranked probability score. We discuss the results based on these measures in the following sections.

## 6.1 Predictive accuracy: Probability integral transform

To assess whether the predictive density is correctly specified, we compute probability integral transforms (PIT), i.e. we evaluate the cumulative density of a given forecast up to the realised value.

$$PIT_{m,h,t} = \int_{-\infty}^{y_t} \phi_t(\hat{y}_{m,t|t-h})d\hat{y}_{m,t|t-h} \equiv \Phi_t(\hat{y}_{m,t|t-h})$$

A PIT of 0 indicates that, in advance, no probability was assigned to the possibility that growth could be lower than the realised value of the target variable. If the PIT has the maximum value of 1, then all the predictive density underestimated the realisation. For any well-behaved density forecast, the PIT should be uniformly distributed between 0 and 1 (Diebold et al., 1998). On average over time, the probability that the realised value is lower than the forecasted value should be the same no matter whether we consider high or low realisations. Figure 5 shows the empirical cumulative distribution of GDP PITs against the theoretical uniform distribution and its confidence interval.

If they followed a uniform distribution, their empirical cumulative distribution function (CDF) would follow the 45-degree line. To test this formally, we apply an augmented version of the Kolmogorov-Smirnov test for uniformity, which accounts for the fact that model parameters are estimated on a finite sample, as proposed by Rossi and Sekhposyan

FIGURE 5: EMPIRICAL CUMULATIVE DISTRIBUTION OF PIT VS UNIFORM
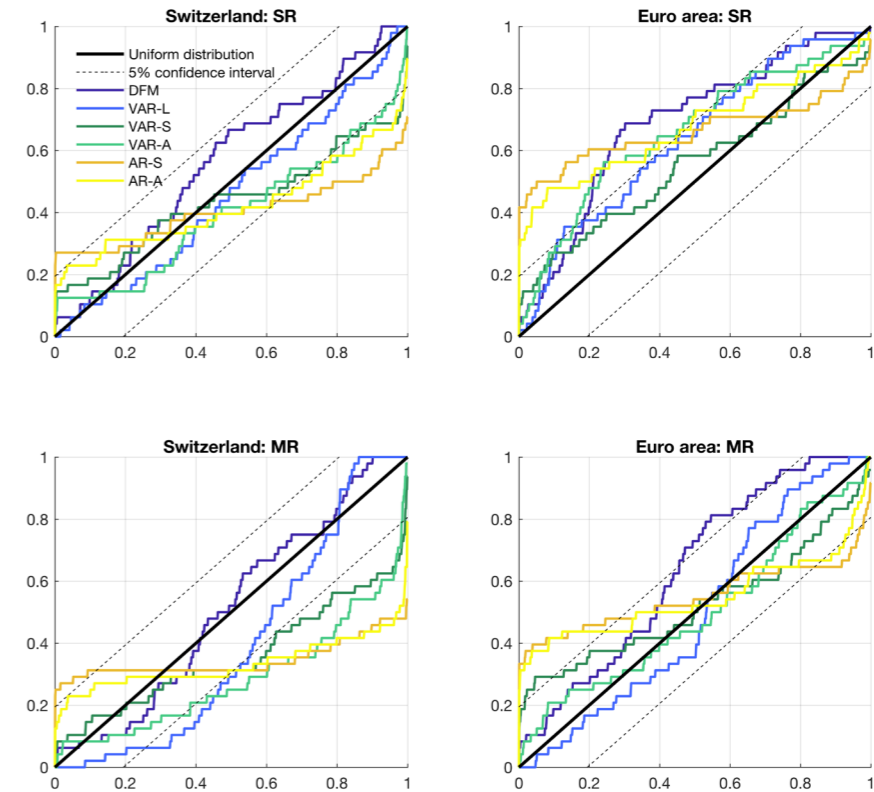
Note: PITs for ordered realised values are on the horizontal axis, the corresponding distribution of the realised value (i.e. the fraction of draws that is lower than the corresponding realised value) on the vertical axis.

(2013). Based on a fine grid $r \in [0,1]$ we calculate

$$\xi_{m,t|t-h}(r) \equiv (1\{\Phi_t(\hat{y}_{m,t|t-h}) \leq r\} - r\})$$

for every grid point.[16] For low values of r, the indicator is typically zero and thus $\xi(r)$ is negative but small. For r in the region of a half, dispersion of the $\xi(r)$ vector is highest as some values are close to 0.5 and some close to -0.5. And for values close to 1, the indicator function is usually 1, and thus $\xi(r)$ positive and small. For every grid point, we calculate

---

[16]Conveniently, one can set the grid r so as to put a special emphasis on parts of the distribution which are of particular interest, such as lower and/or upper tails.

the objective, whose absolute value we maximise

$$\kappa_{KS} = \sup_r \left| \frac{1}{\sqrt{T}} \sum_t^T \xi_{m,t|t-h}(r) \right|$$

The resulting κ_{KS} is evaluated against critical values obtained from a simulation: In a large number of Monte Carlo replications, we draw T random variables from the uniform distribution, calculate κ and use the $(1 - \alpha)$-th percentile of all simulations as the critical value for the α significance level. If $\kappa_{KS} > \kappa_\alpha$, then the test rejects that the empirical distribution could be the result of a uniform data-generating process at the α percent significance level. The corresponding p-values are reported in Table 7.

The two left-hand side panels of Figure 5 show the CDFs for Switzerland. The DFM narrowly follows the pattern of the uniform distribution for most of the distribution, both for the short run and the medium run. The test of uniformity for the VAR-L cannot be rejected, although there are some deviations at the lower end of the distribution for the medium-run CDF. The rather convex CDF of the benchmark VARs (in green) indicate the opposite: too often, the PITs are at the very high end, indicating that the models significantly underestimate the probability of high growth rates. The univariate models, both on the aggregate and sectoral level, have an inverted S-shape. This pattern is especially pronounced for the medium run and suggests that uncertainty has been underestimated with these models.

For the euro area, the CDFs show that the PITs are overall less uniformly distributed than for Switzerland. By inspection, the VARs perform better than the DFM and, of all models under consideration, this is the one that most closely follows the uniform distribution. However, for the VAR-S only, the null of a uniform distribution cannot be rejected at the 5% level. Using euro area data, the DFM tends to overestimate the realised values. Over the entire range, the realised probability is lower than the model implied. However, this misalignment is also visible for the univariate benchmarks. Overall, density forecasts for the euro area perform worse than for Switzerland. Again, this may be related to the fact that the estimation sample is quite short in this case.

6.2 Relative performance: Ranked probability score

When comparing the predictive densities across models, scoring rules derived from the concept of PIT are helpful tools. Various scoring rules such as loss functions may help evaluate models against alternatives (Giacomini and White (2006), Kenny et al. (2012), Boero et al. (2011)). We separate the argument space of the probability density into mutually exclusive events, which can be thought of as bins $k = (1, 2, ..., K)$ in the predictive density of the forecast. We use $K = 16$ intervals set according to the Survey of Professional Forecasters.[17] Every bin is assigned a probability from the distribution, for example for the first bin $\psi_{m,k,t} = \int_{-\infty}^{v(k=1)} \phi(\hat{y}_{m,t|t-h}) d\hat{y}_{m,t|t-h}$. Additionally, we define a vector of length K with binary values: 1 if the realised value is within the respective bin and 0 otherwise $(d_{m,1,t}, d_{m,2,t}, ..., d_{m,K,t})$. Then the inherently Bayesian predictive likelihood score (log score) would be defined as follows:

$$S_{m,h} = \frac{1}{T} \sum_t^T S_{m,h,t}, \qquad S_{m,h,t} = \sum_k^K d_{m,k,t} \log(\psi_{m,k,t})$$

A problem with the log score arose as the specific bin of realisations was assigned a probability of zero such that the log of zero would be negative infinity for a substantial fraction of forecasts for the smaller benchmark models. A possible fix would imply reducing the number of bins to make sure every bin carries positive probability, but this would ultimately violate the purpose and spirit of the exercise.

We therefore use the ranked probability score (RPS, or Epstein score) as the alternative, which is a measure of the cumulative probabilities and indicators.

The RPS is defined as follows:

$$\Psi_{m,k,t} = \sum_{j=1}^k \psi_{m,j,t}, \qquad D_{m,k,t} = \sum_{j=1}^k d_{m,j,t}$$
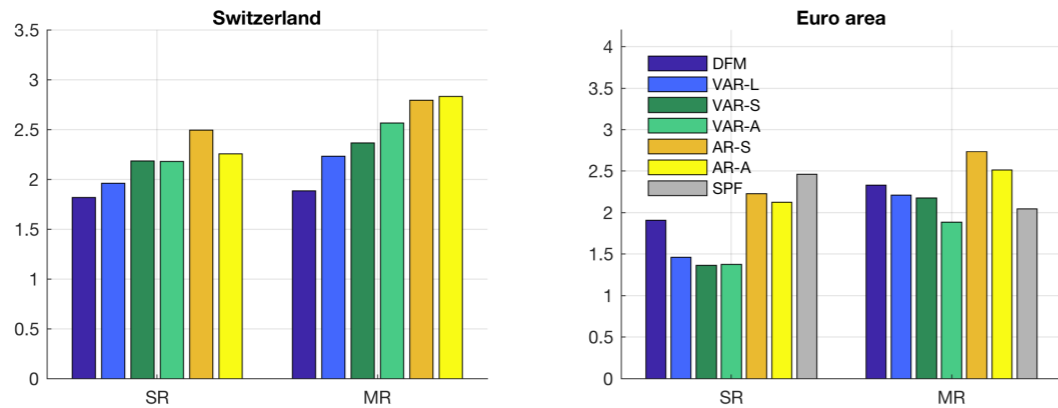
$$RPS_{m,h} = \frac{1}{T} \sum_t^T RPS_{m,h,t}, \qquad RPS_{m,h,t} = \sum_k^K (\Psi_{m,k,t} - D_{m,k,t})^2$$

---

[17]The partitioning of half a percentage point is used on a grid between annual growth rates from -3 to 4 percent.

As values are now in the positive range, it is desirable to have them as small as possible.[18]

Figure 6 summarises the results, which can be found in Table 7 in detail.

FIGURE 6: RANKED PROBABILITY SCORE COMPARED

When we perform a test analogous to Diebold-Mariano in the case of point forecasts, but use the RPS as a loss function (cf. Boero et al. (2011)), the estimated coefficient should be negative in order to beat the basic AR forecast.

$$RPS_{m,h,t} - RPS_{AR\text{-}A,h,t} = \beta_{RPS} + u_t, \qquad H_0 : \beta_{RPS} = 0$$

In the Swiss case, the RPS of the DFM is lower than that of the other models, and the VAR-L finishes in second place. The factor structure helps to improve the performance significantly, especially in the medium run. Among the benchmark models, which all trail the models that include sectoral interlinkages both for the short and medium-run evaluation, the more complex ones outperform the simplest ones. We conclude that allowing for more complex model dynamics as well as interlinkages between sectoral variables tends to improve the performance of medium-run density forecasts.[19]

For the euro area, the relative ranking between the large models is different: The VAR

---

[18]This measure is a discrete approximation to the measure discussed, e.g., in Gneiting and Raftery (2007). Evaluating the continuous version using expression (21) in their paper yields similar results, see bottom row in Table 7. One difference is that the VAR-L and the DFM perform somewhat worse in the euro area, but the difference is quite small against the backdrop of the estimation uncertainty.

[19]In parallel to the evaluation of the point forecasts, we may decompose the density of the forecast error into a marginal and a dependence component using copulas, see e.g. C. Diks and van Dijk (2010). This is left for future investigation.

TABLE 7: DENSITY FORECAST PERFORMANCE

		DFM	VAR-L	VAR-S	VAR-A	AR-S	AR-A
Switzerland							
Mean PIT	SR	0.51	0.59	0.56	0.67	0.60	0.65
	MR	0.51	0.57	0.64	0.66	0.59	0.63
KS/RS p-value	SR	0.31	0.00	0.00	0.00	0.00	0.00
	MR	0.30	0.02	0.00	0.00	0.00	0.00
RPS	SR	1.82	1.96	2.19	2.18	2.49	2.26
	MR	1.89	2.24	2.37	2.57	2.79	2.84
DM test: β_{RPS}	SR	-0.44	-0.30	-0.07	-0.08	0.24	-
		(0.27)	(0.26)	(0.29)	(0.27)	(0.15)	-
	MR	-0.95	-0.60	-0.47	-0.27	-0.04	-
		(0.37)	(0.31)	(0.20)	(0.27)	(0.05)	-
Euro area							
Mean PIT	SR	0.40	0.51	0.45	0.50	0.44	0.50
	MR	0.42	0.50	0.48	0.46	0.41	0.49
KS/RS p-value	SR	0.00	0.40	0.10	0.38	0.00	0.00
	MR	0.00	0.16	0.01	0.07	0.00	0.00
RPS	SR	1.91	1.46	1.37	1.37	2.23	2.13
	MR	2.33	2.21	2.17	1.89	2.74	2.52
DM test: β_{RPS}	SR	-0.22	-0.67	-0.76	-0.75	0.10	-
		(0.32)	(0.30)	(0.33)	(0.31)	(0.10)	-
	MR	-0.19	-0.31	-0.34	-0.63	0.22	-
		(0.21)	(0.46)	(0.24)	(0.30)	(0.13)	-

Note: The mean probability integral transform (PIT) of an unbiased density forecast would be 0.5. The Kolmogorov-Smirnov/Rossi-Sekhposyan (KS/RS) test rejects uniformity at the p-significance level. The log score has a negative scale with maximum 0. The ranked probability score (RPS) has positive support and an inverted scale, i.e. optimum 0, and the respective Diebold-Mariano (DM) test coefficient is negative if the respective model beats the benchmark model AR-A (Newey-West standard errors in brackets).

models without factor structure beat DFM indicating that the value added of sectoral information in density forecasting is limited if sectoral comovement is high. In the short run, the predictive densities are all better than those involving judgment by SPF participants, despite the latter's advantage due to publication lag. The density forecasts for the best performing models even beat SPF in the medium run.

7 Sectoral heterogeneity

Having shown that jointly modelling the dynamics of sectoral production improves point and density forecasts for aggregate GDP, especially for the medium term, we now analyse the cross-sectional forecast distribution. For this evaluation, we rely on standard measures

for multivariate forecasts such as the log determinant of the (weighted) error covariance matrix and a multivariate mean squared error in the following section. We also propose two new measures comparing specific aspects of the forecast distribution.

7.1 Multivariate root mean squared error

The cross-sectional forecast distribution can be evaluated using standard measures for the evaluation of multivariate densities. We follow Carriero et al. (2011) and calculate the multivariate mean squared error as

$$\text{Multi-MSE}_{m,h} = \text{trace}\left(\frac{1}{T}\sum_{t=1}^{T}(e_{m,h,t}^{S})' M^{-1}(e_{m,h,t}^{S})\right)$$

where e_h^S is the matrix of h-step ahead forecast errors of all sectors over time and M a weighting matrix with dimensions SxS containing the variances of the sectoral target series along its diagonal.[20]

The results are shown in the first lines of Table 8.[21] In the Swiss case, we find that simple univariate models tend to outperform those that model comovement explicitly. The result is especially pronounced for the VAR with all sectors and macro variables, while the DFM is more competitive. In the euro area, the DFM is almost as competitive as the benchmark VARs. Overall, the results suggest that there is a potential gain from explicitly modelling sectors and their interactions, although simple models may be preferred to complicated ones in some cases.

This result stands in some contradiction to the evaluation of the (root) MSE in section 5.1 only at first sight. Indeed, the MSE for aggregate GDP and the multivariate MSE calculated here are closely connected in the sense that both measures are weighted sums

---

[20]We use the variances of the 8-quarter growth rates for the short run as well as for the medium run, in order to obtain more stable results.

[21]An alternative measure, Bayesian in spirit, considers the natural logarithm of the determinant of the weighted error covariance matrix:

$$\text{Log determinant}_{m,h} = \ln\left|\frac{1}{T}\sum_{t=1}^{T}\left(M^{-1/2}e_{m,h,t}^{S}\right)\left(M^{-1/2}e_{m,h,t}^{S}\right)'\right|$$

The results are also reported in Table 8. Overall, this measure confirms the results from the multivariate mean squared error.

of sectoral forecast errors. Specifically, the MSE for aggregate GDP is a version of the multivariate MSE, but with a time-varying weighting matrix $M_t = \omega_{t-1}\omega_{t-1}'$ where ω_t is the vector of nominal shares of the sectors in aggregate GDP.[22] Two differences from the multivariate MSE as specified above appear. First, the weights on the diagonal are proportional to the squared weight in aggregate GDP. They thus represent the importance of the sector and not the unconditional variances of the sectors. Secondly, the off-diagonal elements are not zero, but represent the product of the respective sectoral share in the MSE for GDP. So the covariances of the sectoral errors are taken into account according to their weights in aggregate GDP in the RMSE. This is not the case in the standard implementation of the MSE with a diagonal weighting matrix M. Interestingly, given the results from the decomposition in section 5.2, we may even expect that the multivariate MSE does not favour the DFM over the simple models, as the gain in forecast performance mainly stems from a better description of the sectoral covariances. Taken together, we think that the evaluation of the RMSE for aggregate GDP is a better summary of the sectoral forecast performance than the multivariate MSE as specified in this section.

7.2 Alternative measures for sectoral developments

The measures analysed so far aim at providing an encompassing assessment of the multivariate forecast distribution. This comes with the drawback that it is difficult to know where the differences in forecasting performance stem from. We therefore now evaluate the forecasting performance based on two additional criteria geared towards capturing other relevant aspects of the cross-sectional forecast distribution.

The first measure compares the weighted share of sectors that were correctly projected to grow above or below their long-term average respectively.[23] This reflects the idea that a model is useful if it is able to tell in which direction the specific sectors are going to develop.

---

[22]This can be seen as follows:

$$\frac{1}{T}\sum_{t=1}^{T}e_{m,h,t}^{2} = \frac{1}{T}\sum_{t=1}^{T}(\omega_{t-1}'e_{m,h,t}^{S})^{2} = \frac{1}{T}\sum_{t=1}^{T}(\omega_{t-1}'e_{m,h,t}^{S})'(\omega_{t-1}'e_{m,h,t}^{S}) = \frac{1}{T}\sum_{t=1}^{T}e_{m,h,t}^{S'}\omega_{t-1}\omega_{t-1}'e_{m,h,t}^{S}$$

[23]As discussed in the previous section, we think that it is important to take the importance of the sector for aggregate GDP into account, which is why we have calculated the weighted share.

The results are shown in Table 8. Using Swiss data, the weighted share of sectors correctly predicted for the first year forecasted was 58% for the DFM, compared to 46% for the aggregate AR that assumes all sectors grow the same. Adding sectoral components improves the share predominantly in the short run (both in Switzerland and the euro area) and the factor model, while best performing in all cases, adds most of the value in the medium run.

The second measure assesses how well models predict the dispersion of growth across the economy as measured by its cross-sectoral standard deviation. The idea behind this measure is that a model can be useful if it correctly predicts how different the sectors are from each other, independent of its sectoral point forecast performance.

We construct the measure as follows. Using the draws from the posterior distribution and draws from the implied distribution of the error terms, we draw from the distribution of the sectoral series over the forecast sample. For each forecast horizon, we take the mean of the cross-sectoral standard deviation $\hat{\sigma}_s$ across draws and compare it to the realised cross-sectoral dispersion in the corresponding time period (σ_s). The root mean squared dispersion error is then defined as

$$RMSDE_{m,h} = \sqrt{\frac{1}{T}\sum_t^T (\hat{\sigma}_{s,t,m,h} - \sigma_{s,t})^2}, \qquad \hat{\sigma}_{s,t,m,h} = \sqrt{\frac{1}{S}\sum_{s=1}^S (\hat{x}_{t,m,h}^s - \frac{1}{S}\sum_{s=1}^S \hat{x}_{t,m,h}^s)^2}$$

The results in Table 8 show that, for this measure, the DFM yields the best performance in the euro area, while it is comparable with the simple benchmarks in Switzerland. However, it is quite striking that the VAR-L performs worse than the other models in both regions. This suggests that treating macro variables and sectoral variables symmetrically as in the VAR-L is probably too crude a way of shrinking the parameter space.

		DFM	VAR-L	VAR-S	VAR-A	AR-S	AR-A
Switzerland							
Multivariate error measures							
Log determinant	SR	-0.32	3.53	-2.20	-4.27	-3.93	-5.05
	MR	-3.62	-1.98	-4.33	-5.52	-5.58	-6.60
Multivariate MSE	SR	20.68	27.05	22.04	21.16	16.14	18.04
	MR	17.14	23.36	20.70	18.84	16.52	16.10
Sectoral dispersion							
Above/below average	SR	0.58	0.53	0.55	0.48	0.55	0.46
	MR	0.55	0.51	0.52	0.52	0.51	0.51
RMS dispersion error	SR	2.94	3.62	2.08	3.97	1.93	3.97
	MR	3.68	8.87	2.56	3.97	2.47	3.97
Euro area							
Multivariate error measures							
Log determinant	SR	-7.03	-1.32	-4.68	-7.68	-2.78	-6.79
	MR	-8.56	-1.37	-7.90	-8.64	-3.28	-6.27
Multivariate MSE	SR	15.16	27.98	18.39	41.84	21.20	66.37
	MR	16.01	34.45	16.58	17.42	33.05	84.79
Sectoral dispersion							
Above/below average	SR	0.65	0.59	0.61	0.49	0.60	0.50
	MR	0.60	0.53	0.55	0.50	0.57	0.50
RMS dispersion error	SR	1.48	3.66	2.52	2.21	3.00	2.21
	MR	1.95	11.28	4.48	2.21	8.34	2.21

8 Conclusions

Against the backdrop of an unprecedented deep recession and the extraordinary conduct of monetary policy over the last decade, the analysis of sectoral heterogeneity in relation to macroeconomic developments has sparked strong interest among policy makers and academics. With our analysis, we contribute to the literature by analysing the forecasting performance of models that are particularly suitable in this respect. Specifically, we measure the forecasting performance of different models based on the real value added series of the different production sectors of the economy. In our evaluation, we focus on medium-term projections for GDP.[24]

We find quite distinct evidence that a factor model structure performs very well. It very

---

[24]Equivalent measures for point and density forecasts of CPI/HICP inflation can be found in the appendix 9.3. The RMSE gain is between 21 (Switzerland) and 29 percent (euro area). Differences among sectoral and aggregate benchmarks are only marginal, but note that we do not include prices on a sectoral level. The Dynamic Factor Model performs very competitively. Density forecasts of inflation do not improve substantially.

generally outperforms the simple benchmarks, and in many cases also the BVAR. This is true for both point and density forecasts. In the latter case, the differences tend to be even more pronounced. Our analysis of the covariance matrix of the sectoral forecast errors suggests that the superiority of the factor model can be traced back to its ability to capture sectoral comovement more accurately than its competitors. Moreover, the factor model tends to outperform the other models also at forecasting sectoral heterogeneity. In particular, it forecasts more accurately the sectoral dispersion as measured by the cross-sectional standard deviation of the sectors.

Thus production-side models, and especially the DFM, provide a valuable complement to demand-side, medium-term models. This is particularly because they allow us to study how different sectors behave in alternative macroeconomic scenarios.

References

ANDERSON, B. D. O. AND J. B. MOORE (1979): *Optimal Filtering*, Prentice Hall.

BAI, J. AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

BANBURA, M., D. GIANNONE, AND L. REICHLIN (2010): "Large Bayesian Vector Autoregressions," *Journal of Applied Econometrics*, 25, 71–92.

BARHOUMI, K., O. DARNÉ, L. FERRARA, AND B. PLUYAUD (2012): "Monthly GDP forecasting using bridge models: Application for the French economy." *Bulletin of Economic Research*, 64, 53–70.

BAUMEISTER, C. AND L. BENATI (2013): "Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound," *International Journal of Central Banking*, 9, 165–212.

BÄURLE, G. (2013): "Structural Dynamic Factor Analysis Using Prior Information From Macroeconomic Theory," *Journal of Business and Economic Statistics*, 31, 136–150.

BÄURLE, G. AND E. STEINER (2015): "How Do Individual Sectors Respond to Macroeconomic Shocks? A Structural Dynamic Factor Approach Applied to Swiss Data," *Swiss Journal of Economics and Statistics*, 151, 167–225.

BAUWENS, L., M. LUBRANO, AND J.-F. RICHARD (1999): *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press.

BERMINGHAM, C. AND A. D'AGOSTINO (2011): "Understanding and Forecasting Aggregate and Disaggregate Price Dynamics," *European Central Bank Working Paper Series*, 1365.

BERNANKE, B., J. BOIVIN, AND P. S. ELIASZ (2005): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) approach," *The Quarterly Journal of Economics*, 120, 387–422.

BERNHARD, S. (2016): "A real-time GDP data set for Switzerland," Tech. rep., Swiss National Bank Economic Studies 2016-09.

BOERO, G., J. SMITH, AND K. F. WALLIS (2011): "Scoring Rules and Survey Density Forecasts," *International Journal of Forecasting*, 27, 379–393.

BOIVIN, J. AND M. GIANNONI (2006): "DSGE Models in a Data-Rich Environment," NBER Technical Working Paper 0332, National Bureau of Economic Research, Inc.

BOIVIN, J. AND S. NG (2006): "Are More Data Always Better for Factor Analysis?" *Journal of Econometrics*, 132, 169–194.

C. DIKS, V. P. AND D. VAN DIJK (2010): "Out-of-sample comparison of copula specifications in multivariate density forecasts," *Journal of Economic Dynamics and Control*, 9, 1596–1609.

CARRIERO, A., T. CLARK, AND M. MARCELLINO (2014): "No Arbitrage Priors, Drifting Volatilities, and the Term Structure of Interest Rates," *CEPR Discussion Papers*, 9848.

CARRIERO, A., G. KAPETANIOS, AND M. MARCELLINO (2011): "Forecasting Large Datasets with Bayesian Reduced Rank Multivariate Models," *Journal of Applied Econometrics*, 26, 735–761.

CARTER, C. K. AND P. KOHN (1994): "On Gibbs Sampling for State Space Models," *Biometrica*, 81, 541–553.

CARVALHO, V. M., M. NIREI, Y. U. SAITO, AND A. TAHBAZ-SALEHI (2016): "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake," *CEPR Discussion Papers*, 11711.

CHANG, Y. AND S. HWANG (2015): "Asymmetric Phase Shifts in U.S. Industrial Production Cycles," *Review of Economics and Statistics*, 97, 116–133.

CHIB, S. (1993): "Bayes regression with autoregressive errors : A Gibbs sampling approach," *Journal of Econometrics*, 58, 275–294.

CHRISTIANO, L. AND T. FITZGERALD (1998): "The business cycle: it's still a puzzle," *Economic Perspectives*, 22, 56–83.

DÉES, S. AND J. GÜNTHER (2014): "Analysing and Forecasting Price Dynamics Across Euro Area Countries and Sectors: A Panel VAR Approach," *European Central Banking Working Paper Series*, 1724.

DIEBOLD, F. X., T. A. GUNTHER, AND A. S. TAY (1998): "Evaluating Density Forecasts with Application to Financial Risk Management," *International Economic Review*, 39, 863–883.

DIEBOLD, F. X. AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

DOAN, T., R. B. LITTERMAN, AND C. A. SIMS (1984): "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100.

DRECHSEL, K. AND R. SCHEUFELE (2012): "Bottom-up or Direct? Forecasting German GDP in a Data-rich Environment," *Swiss National Bank Working Papers*, 16.

FARES, J. AND G. SROUR (2001): "The Monetary Transmission Mechanism at the Sectoral Level," Working Paper 2001-27, Bank of Canada.

FRÜHWIRTH-SCHNATTER, S. (1994): "Data augmentation in dynamic linear models," *Journal of Time Series Analysis*, 15, 183–202.

FUENTES-ALBERO, C. AND L. MELOSI (2013): "Methods for computing marginal data densities from the Gibbs output," *Journal of Econometrics*, 175, 132 – 141.

GANLEY, J. AND C. SALMON (1997): "The Industrial Impact of Monetary Policy Shocks: Some Stylized Facts," Working Paper 68, Bank of England.

GEWEKE, J. (2005): *Contemporary Bayesian Econometrics and Statistics*, John Wiley & Sons, Inc.

GIACOMINI, R. AND H. WHITE (2006): "Tests of Conditional Predictive Ability," *Econometrica*, 74, 1545–1578.

GIANNONE, DOMENICO, M. L. AND G. E. PRIMICERI (2015): "Prior selection for vector autoregressions," *Review of Economics and Statistics*, 97, 436–451.

GNEITING, T. AND A. E. RAFTERY (2007): "Strictly Proper Scoring Rules, Prediction and Estimation," *Journal of the American Statistical Association*, 102, 359–378.

HAHN, E. AND F. SKUDELNY (2008): "Eearly Estimates of Euro Area real GDP Growth. A Bottom Up Approach from the Production Side," *European Central Bank Working Paper Series*, 975.

HENDRY, D. F. AND K. HUBRICH (2011): "Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate," *Journal of Business and Economic Statistics*, 29, 216–227.

HORVATH, M. (1998): "Cyclicality and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks," *Review of Economic Dynamics*.

HUBRICH, K. (2005): "Forecasting Euro Area Inflation: Does Aggregating Forecasts by HICP Component Improve Forecast Accuracy?" *International Journal of Forecasting*, 21, 119–136.

KARLSSON, S. (2013): "Forecasting with Bayesian Vector Autoregressions," in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier, vol. 2, chap. 15, 791–897.

KENNY, G., T. KOSTKA, AND F. MASERA (2012): "How Informative Are the Subjective Density Forecasts of Macroeconomists?" *European Central Bank Working Paper Series*, 1446.

KIM, C.-J. AND C. R. NELSON (1999): *State-Space Models with Regime Switching*, The MIT Press.

LITTERMAN, R. (1979): "Techniques of forecasting using vector autoregressions," *Federal Reserve of Minneapolis Working Paper*.

MARCELLINO, M., J. H. STOCK, AND M. W. WATSON (2003): "Macroeconomic Forecasting in the Euro area: Country Specific Versus Area-wide Information," *European Economic Review*, 47, 1–18.

MARTINSEN, K., F. RAVAZZOLO, AND F. WULFSBERG (2014): "Forecasting Macroeconomic Variables Using Disaggregate Survey Data," *International Journal of Forecasting*, 30, 65–77.

MINCER, J. A. AND V. ZARNOWITZ (1969): *The Evaluation of Economic Forecasts*, National Bureau of Economic Research, New York.

NEUSSER, K. (2009): *Zeitreihenanalyse in den Wirtschaftswissenschaften*, Vieweg + Teubner Verlag, 3rd ed.

NEWEY, W. K. AND K. D. WEST (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

PESARAN, M. H., T. S. AND S. M. WEINER (2010): "Modelling regional interdependencies using a global error-correcting macroeconometric model," *Journal of Business and Economics Statistics*, 129–162.

POIRIER, D. (1998): "Revising Beliefs in Nonidentified Models," *Econometric Theory*, 14, 483–509.

PRIMICERI, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72, 821–52.

ROSSI, B. (2013): "Advances in Forecasting under Model Instability," in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier, vol. 2B, 1203–1324.

ROSSI, B. AND T. SEKHPOSYAN (2013): "Conditional Predictive Density Evaluation in the Presence of Instabilities," *Journal of Econometrics*, 177, 199–212.

STOCK, J. H. AND M. W. WATSON (2002): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, 147–162.

TAYLOR, J. (1978): "Comments on "the temporal and sectoral aggregation of seasonally adjusted time series" by John Geweke," Economic research report, US Department of Commerce, Bureau of the Census.

ZELLNER, A. AND J. TOBIAS (2000): "A Note on Aggregation, Disaggregation and Forecasting Performance," *Journal of Forecasting*, 19, 457–469.

9 Appendix

9.1 Detailed model description and estimation method

We provide some details on the estimation procedure, allowing the reader to replicate our empirical results. Additionally, we provide detailed references to previous work where the formal derivations of the (conditional) posterior distributions can be found.

As the posterior distribution cannot be derived analytically, we use Markov Chain Monte Carlo (MCMC) methods to simulate from the posterior distribution. In our setting, this can be done using a Gibbs sampling approach (see e.g. Kim and Nelson (1999) with one iteration of the Gibbs sampler involving the following steps:

Step 1: Draw the factors conditional on a set of model parameters

Step 2: Draw parameters in the observation equation conditional on the factors

Step 3: Draw parameters in the state equation conditional on the factors

Iterating over these steps delivers draws from the posterior distribution of the parameters and the factors. Subsequently, we provide a detailed description of the three steps including the specification of the prior distribution.

Step 1: Drawing the factors To draw from the joint distribution of the factors given the parameters in the model, we use the algorithm of Carter and Kohn (1994) and Frühwirth-Schnatter (1994). The algorithm uses a Kalman filter. In our setting, the filter has to be adapted for autoregressive errors and potentially co-linear states. See, e.g., Anderson and Moore (1979) and Kim and Nelson (1999).

Step 2: Drawing parameters in the observation equation We use an informative prior on the factor loadings as this 'identifies' the factors in the sense that it puts curvature into the posterior density function for regions in which the likelihood function is flat. See, for example, the discussion in Bäurle (2013). In our implementation, the prior is centred such that, a priori, the series are all related with loading one to the unobserved factors contemporaneously and with loading zero to the lagged factors. However, the variance of

the prior is chosen to be large, such that if the data is informative about the loadings, this will be reflected in the posterior distribution.

Regarding the parametric form of the prior, we use the specification of the conjugate prior described in Bauwens et al. (1999), p.58: The prior distribution $p(R_n, \Lambda_n \mid \Psi_n)$, where n denotes the respective row in the observation equation, is of the normal-inverted gamma-2 form (as defined in the appendix of Bauwens et al. (1999)):

$$R_n \sim \mathrm{iG}_2(s, \nu)$$

$$\Lambda_n \sim \mathrm{N}(\Lambda_{0,n}, R_n M_{0,n}^{-1})$$

Λ_0 is the prior mean of the distribution. The parameters s and ν parametrise the distribution of the variance of the measurement error. M_0 is a matrix of parameters that influences the tightness of the priors in the observation equation. The larger the elements of M_0, the closer we relate the observed series to the factors a priori. The choice of the tightness is determined by the a priori confidence in the prior belief. We set $M_{0,n,\varrho} = \varrho^2$ for all n and $\varrho = 1, \ldots, q$. Thus the tightness of the prior increases quadratically with the lag of the factor. Following Boivin and Giannoni (2006), we set $s = 3$ and $\nu = 0.001$. By adding a standard normal prior for Ψ_n, we have specified a complete prior distribution for the parameters in the observation equation. The derivation of the posterior distribution is standard, see e.g. Chib (1993) and Bauwens et al. (1999).

Step 3: Drawing parameters in the state equation The procedure for drawing from the state equation conditional on the factor is identical to the estimation of the BVAR. We implement a normal Wishart prior for the parameters in the state equation. The prior mean and variances are of a Minnesota type, following Banbura et al. (2010). In that notation, we set the prior mean of the first own lag to zero as we model stationary series. The prior is conjugate, i.e. the conditional densities $p(\Sigma|F, \Phi)$ and $p(\Phi|F, \Sigma)$ can be shown to be multivariate normal and inverse Wishart densities respectively (see Bauwens et al. (1999) or Karlsson (2013)). We therefore introduce this additional Gibbs-sampling step into our MCMC algorithm.

9.2 GDP nowcast evaluation

For completeness, we test if the models are correctly specified using the Mincer-Zarnowitz test. Mincer and Zarnowitz (1969) argue that even if estimated coefficients are unbiased, the resulting forecasts may underestimate high values and overestimate low values. However, if $\beta_0 = 0$ and $\beta_1 = 1$ in the following regression, the forecast is unbiased and high forecasts are followed by equally high realisations in expectation:

$$\tilde{y}_{t,t-h} = \beta_0 + \beta_1 \hat{y}_{t|t-h} + u_t, \qquad H_0 : \beta_0 = 0, \beta_1 = 1$$

Table 9 reports the respective regression coefficients for the one-quarter-ahead prediction. We report Newey-West standard errors in brackets (Newey and West, 1987), which are adjusted for possible heteroskedasticity and autocorrelation.

It also contains other metrics discussed in the article, such as RMSE and RPS, for GDP forecasts one quarter ahead, to make them comparable to other evaluations in the literature.

TABLE 9: EVALUATION MEASURES FOR SHORT-RUN GDP FORECASTS

		DFM	VAR-L	VAR-S	VAR-A	AR-S	AR-A
Switzerland							
β_0	h=1	-0.00	0.05	0.02	0.06	-0.06	0.01
		(0.14)	(0.17)	(0.18)	(0.15)	(0.44)	(0.21)
β_1		0.90	0.68	0.81	0.75	0.93	0.80
		(0.18)	(0.19)	(0.24)	(0.19)	(0.65)	(0.25)
RMSE	h=1	0.50	0.55	0.53	0.52	0.58	0.55
RPS	h=1	0.61	0.64	0.72	0.70	0.85	0.76
Euro area							
β_0	h=1	-0.02	-0.06	-0.01	-0.11	-0.06	-0.04
		(0.14)	(0.16)	(0.14)	(0.16)	(0.18)	(0.17)
β_1		0.66	0.59	0.62	0.66	0.59	0.53
		(0.20)	(0.19)	(0.19)	(0.18)	(0.22)	(0.19)
RMSE	h=1	0.60	0.66	0.62	0.64	0.70	0.73
RPS	h=1	0.57	0.51	0.56	0.55	0.65	0.69

Note: Regression of the GDP point one-step-ahead forecast on realised values with Newey-West standard errors in brackets

9.3 Inflation forecast evaluation

All multivariate models can be used to produce forecasts of CPI (Switzerland) and HICP (euro area) inflation. Instead of the AR-S and AR-A benchmark, we run a simple univariate AR of inflation. The following table contains a small subset of the results analogous to GDP in sections 5 and 6.

		DFM	VAR-L	VAR-S	VAR-A	AR	SPF
Switzerland							
RMSE	SR	1.52	1.89	1.97	1.92	1.89	
	MR	1.46	1.75	1.85	1.84	1.73	
RPS	SR	1.42	1.81	2.74	2.31	2.58	
	MR	1.48	1.84	2.57	2.34	2.43	
Euro area							
RMSE	SR	1.26	1.83	1.47	1.40	1.73	1.00
	MR	1.32	1.80	1.38	1.24	1.50	1.12
RPS	SR	1.34	1.75	1.92	1.66	1.86	1.52
	MR	1.46	1.93	1.93	1.75	1.92	1.40

9.4 Stability over time

To assess dynamic performance, we implement the estimate of the local RMSE as suggested by Rossi (2013). We set the window to $m = 16$, i.e. we compute that RMSE over 4 years. This is lower than the values used in Rossi (2013) (m=25 and m=60). Our choice, dictated by the short sample, reduces the local bias but decreases efficiency. While no distinct patterns appear in a visual inspection of these measures (Figure 7), the cumulative errors in Figure 8 reveal that the DFM performs better than the other models, especially in the first half of the sample.
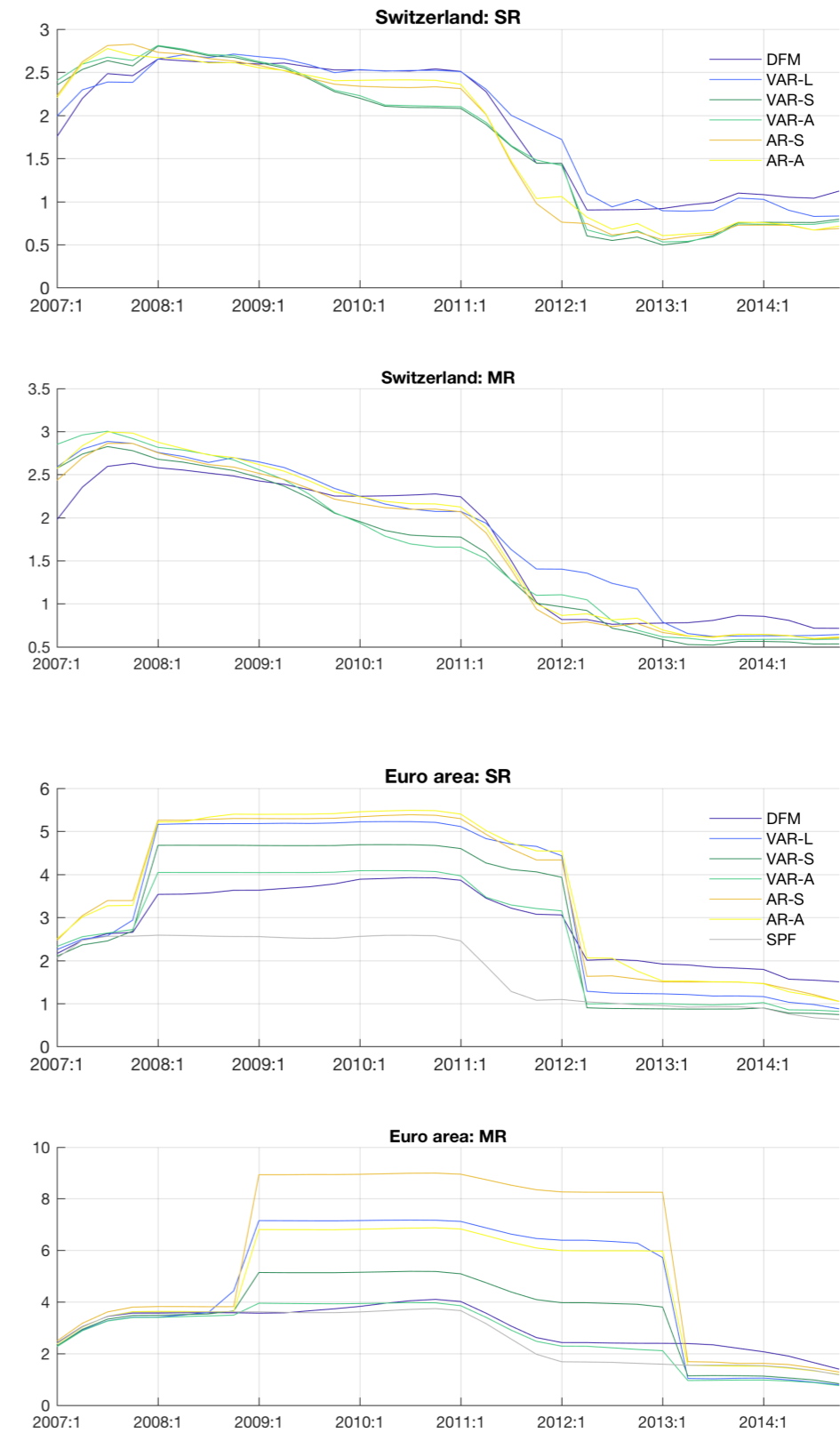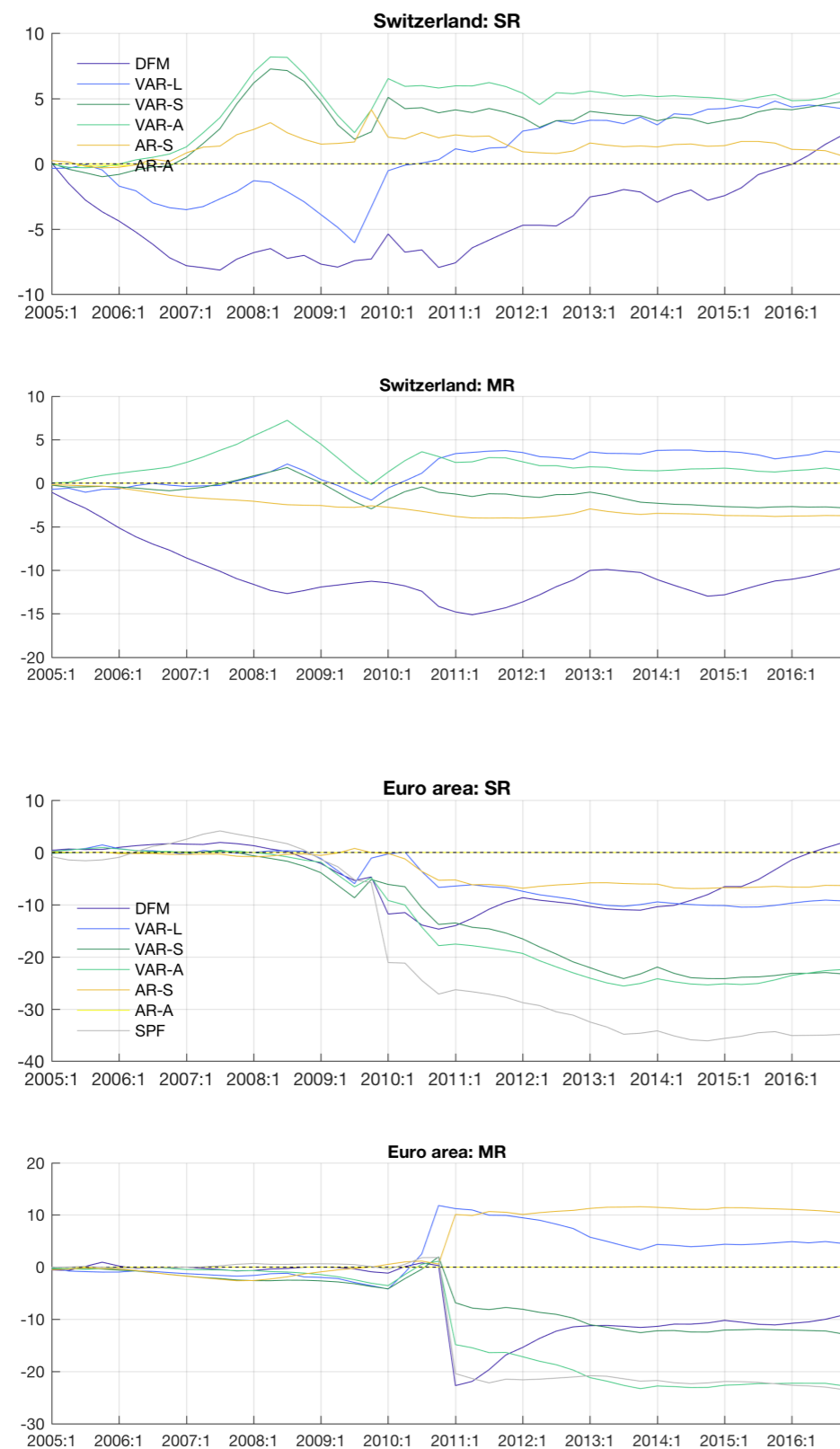
Figure 7: 16-quarter rolling windows of RMSE

FIGURE 8: CUMULATED RELATIVE ERROR

Switzerland: SR

Switzerland: MR

Euro area: SR

Euro area: MR

Recent SNB Working Papers

2018-16 Gregor Bäurle, Elizabeth Steiner and Gabriel Züllig:
Forecasting the production side of GDP

2018-15 Christian Grisse, Gisle J. Natvik:
Sovereign debt crises and cross-country assistance

2018-14 Thomas Lustenberger:
Has the American Output Growth Path Experienced
a Permanent Change?

2018-13 Stephan Imhof, Cyril Monnet and
Shengxing Zhang:
The Risk-Taking Channel of Liquidity Regulations
and Monetary Policy

2018-12 Andreas M. Fischer, Henrike Groeger,
Philip Sauré and Pinar Yeşin:
Current account adjustment and retained earnings

2018-11 Maxime Phillot, Rina Rosenblatt-Wisch:
Inflation Expectations: The Effect of Question
Ordering on Forecast Inconsistencies

2018-10 Alex Cukierman, Thomas Lustenberger:
International Evidence on Professional
Interest Rate Forecasts:
The Impact of Forecasting Ability

2018-9 Thomas Nitschka:
Did China's anti-corruption campaign affect
the risk premium on stocks of global luxury
goods firms?

2018-8 Andreas M. Fischer, Lucca Zachmann:
Do the rich pay their taxes early?

2018-7 Basil Guggenheim, Mario Meichle and Thomas Nellen:
Confederation debt management since 1970

2018-6 Lucas Marc Fuhrer, Basil Guggenheim
and Matthias Jüttner:
What do Swiss franc Libor futures really tell us?

2018-5 Barthélémy Bonadio, Andreas M. Fischer
and Philip Sauré:
The speed of exchange rate pass-through

2018-4 Simone Auer, Christian Friedrich, Maja Ganarin,
Teodora Paligorova, and Pascal Towbin:
International Monetary Policy Transmission trough
Banks in Small Open Economies

2018-3 Romain Baeriswyl, Camille Cornand and Bruno Ziliotto:
Observing and shaping the market: the dilemma of
central banks

2018-2 Adriel Jost:
Cultural Differences in Monetary Policy Preferences

2018-1 Reto Foellmi, Adrian Jaeggi and Rina Rosenblatt-Wisch:
Loss Aversion at the Aggregate Level Across Countries
and its Relation to Economic Fundamentals

2017-18 Gregor Bäurle, Sarah M. Lein and Elizabeth Steiner:
Employment Adjustment and Financial Constraints –
Evidence from Firm-level Data

2017-17 Thomas Lustenberger and Enzo Rossi:
The Social Value of Information:
A Test of a Beauty and Non-Beauty Contest

2017-16 Aleksander Berentsen and Benjamin Müller:
A Tale of Fire-Sales and Liquidity Hoarding

2017-15 Adriel Jost:
Is Monetary Policy Too Complex for the Public?
Evidence from the UK
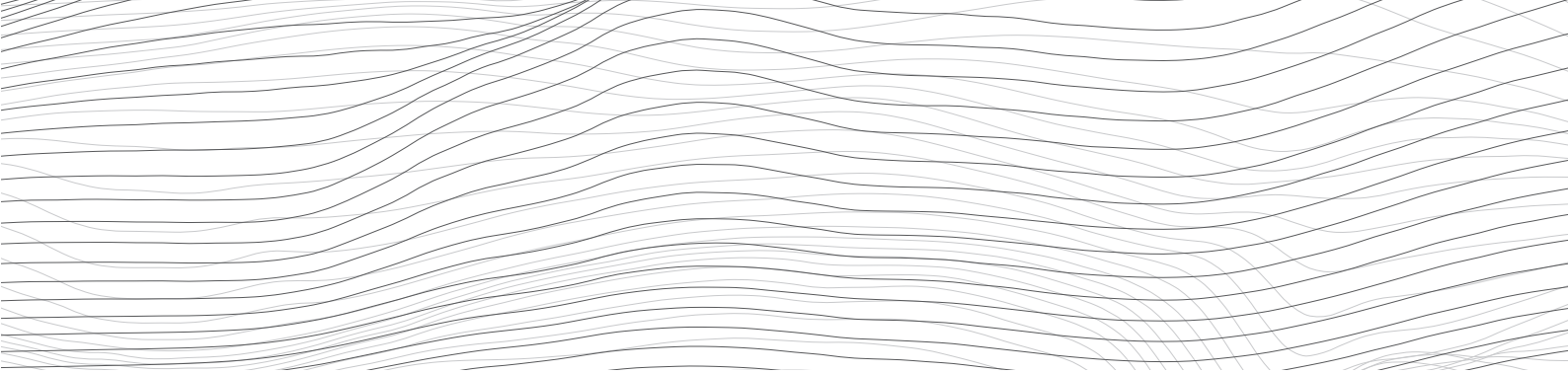
SCHWEIZERISCHE NATIONALBANK
BANQUE NATIONALE SUISSE
BANCA NAZIONALE SVIZZERA
BANCA NAZIUNALA SVIZRA
SWISS NATIONAL BANK   ⊹