Mixed-Frequency Models for Tracking Short-Term Economic Developments in Switzerland*

Alain Galli, Christian Hepenstrick, and Rolf Scheufele
Swiss National Bank

We compare several methods for monitoring short-term economic developments in Switzerland. Based on a large mixed-frequency data set, the following approaches are presented and discussed: a factor-based *information combination* approach (a dynamic factor model based on the Kalman filter/smoother and estimated by the EM algorithm), a *model combination* approach resting on MIDAS regression models, and a *variable selection* approach using a specific-to-general algorithm. In an out-of-sample GDP forecasting exercise, we show that the considered approaches clearly beat relevant benchmarks such as univariate time-series models and models that work with just one indicator. Moreover, we find that the factor model is superior to the other two approaches under investigation. However, forecast pooling of the three methods turns out to be even more promising.

JEL Codes: C32, C53, E37.

*We want to thank Luca Moretti, Massimiliano Marcellino, Lucrezia Reichlin, and three anonymous referees as well as participants at IAAE conference 2016 in Milano, the ISF 2016 in Santander, and the CIRET conference 2016 in Copenhagen for valuable comments. The views, opinions, findings, and conclusions or recommendations expressed in this paper are strictly those of the authors. They do not necessarily reflect the views of the Swiss National Bank. The SNB takes no responsibility for any errors or omissions in, or for the correctness of, the information contained in this paper. In particular, the models described in the paper are not part of the suite of models used for Swiss National Bank's inflation forecast, published on a quarterly basis in its Quarterly Bulletin. Authors' e-mails: alain.galli@snb.ch, christian.hepenstrick@snb.ch, and rolf.scheufele@snb.ch.

1. Introduction

Policy institutions such as central banks depend on the timely assessment of past, current, and future economic conditions. However, key statistics on economic conditions are often estimated infrequently and published with a substantial time delay. GDP, for example, is computed at the quarterly frequency. Looking across countries, the timeliest first estimates are published about one month after the reference quarter (e.g., in the United States, the euro area, the United Kingdom, and China), whereas for Switzerland, the publication lag of GDP is two months. Therefore, the problem of imperfect and incomplete information on the current state of the economy is more pronounced than for other developed economies: many of the main indicators that economists routinely use to track real activity in economies such as the United States, the euro area, or its individual member countries are—if available at all—released only with a long lag, at a quarterly frequency, and/or with very short histories. Still, if one considers all the economic series that are available, it is possible to construct quite a large data set. While each individual indicator of this data set may not look too informative on its own, the data set as a whole looks quite promising, as it features indicators on all important aspects of the Swiss economy.

In this paper, we compare several methods for tracking current economic conditions in Switzerland using this large data set. The two common features of the methods is that they are able to handle (i) large data sets and (ii) indicators with different frequencies. They can therefore be characterized as large-scale mixed-frequency methods. The study relates to similar investigations that have been performed for other countries. Evans (2005), Giannone, Reichlin, and Small (2008), Bańbura, Giannone, and Reichlin (2011), Bańbura and Rünstler (2011), and Kuzin, Marcellina, and Schumacher (2011) are some prominent examples.

We consider three approaches: information combination, model combination, and variable selection. For *information combination*, we rely on factor model techniques.[1] Such models assume that a small number of factors can be used to describe the fluctuations in a

[1] Large Bayesian VARs would be an alternative, but so far they have only been used with single frequencies, and adapting them to the use of mixed frequencies would create computational challenges.

given data set and that those factors are highly correlated with the business cycle and GDP (see, e.g., Stock and Watson 2002, 2011). In this study, we concentrate on a large-scale dynamic factor model (DFM), which is based on the Kalman filter and is estimated by the expectation–maximization (EM) algorithm. This is the current state-of-the-art model for nowcasting purposes (see Bańbura, Giannone, and Reichlin 2011, Doz, Giannone, and Reichlin 2012, and Bańbura and Modugno 2014). The second approach uses *model combination*, i.e., it combines forecasts from different models. In our case, single-indicator models are used to explain GDP within a linear regression framework. For higher-frequency variables (weekly and monthly indicators), mixed-data sampling (MIDAS) is used to take into account the temporal aggregation issue. The third approach is *variable selection*. In this case, we employ a variable selection technique based on a specific-to-general methodology that can handle large data sets and mixed-frequency settings. As benchmark models, we consider univariate time-series models as well as popular Swiss leading indicators such as the PMI and the KOF economic barometer.

To compare the models under investigation and their benchmarks, we conduct a pseudo-out-of-sample forecast exercise. Using our large data set of more than 600 variables at weekly, monthly, and quarterly frequency, and taking into account two different states of information, we generate forecasts from the different models for the period 2005:Q1–2015:Q2.[2] With all our approaches, we take into account the problem of missing observations at the end of the sample (ragged-edge problem). We compare the forecasting performances by means of root mean squared forecast errors (RMSEs) and investigate their ranking over time by looking at different subperiods and rolling windows. Besides RMSEs, we investigate the models' ability to forecast business cycle phases.

Our results show that all three approaches using the large data set clearly beat relevant benchmarks such as univariate time-series models and the popular single-indicator models using the PMI or the KOF economic barometer. The factor model performs best for

[2]The reason for not having a longer evaluation sample is that for the case of Switzerland, many economic indicators start only in 1990 or even later, which automatically limits the estimation sample of the models to some extent.

the total evaluation sample. Model combination also performs well, but it does not capture the financial crisis period very well. After the crisis, model combination slightly outperforms the factor model, particularly for the two-step-ahead forecast. Variable selection performs well on average for the one-step-ahead forecast but loses ground when forecasting two steps ahead. Finally, we show that a pooling (simple average) of the models under consideration produces even better results than the best single-model procedure.

2. Data

To understand the choice of forecasting approaches that we evaluate, some background on the Swiss data situation is warranted. An economist who wants to monitor current activity in Switzerland faces a data situation that is worse than what one may expect given Switzerland's high per-capita income: many indicators that are routinely used to monitor real activity in economies such as the United States and the euro area are released with a considerable lag, are available only at a quarterly frequency, and/or have short histories.

A prominent example that illustrates all three issues is the industrial production, turnover, and new orders release: its publication lag is about two months after the end of the quarter. Moreover, while the time series are available at a monthly frequency, the values for all three months of the quarter are released at once. Finally, the monthly series only go back to 2011, illustrating the problem of short histories; the quarterly series go further back.

Another example of a long publication lag and a quarterly frequency is the job statistics (the equivalent to the firm survey part of the United States' monthly employment situation release), which is published about two months after the end of the quarter. Moreover, it is only available at a quarterly frequency. Other important data releases that are only available at a quarterly frequency include construction activity, wages, and consumer confidence.

A prominent example of a short history is certainly the quarterly labor force survey (the equivalent to the household survey part of the United States' employment situation release) that only goes back to 2010. Still, there exist indicators with short publication lags and monthly frequency. However, they are often quite noisy and/or not too strongly connected with GDP. Examples are merchandise exports and imports, financial markets data, and foreign activity indicators.

In light of this data situation, we think it makes sense to use all available information[3] and construct an as large as possible data set that includes both quarterly and monthly indicators. This novel data set for the Swiss economy consists of 627 variables (361 monthly, 266 quarterly) covering seventeen areas. All data are used in calendar-adjusted, seasonally adjusted, and outlier-adjusted terms.[4] Details of the data set are shown in table 1. In principle, the data set starts in 1975. However, many indicators start only in 1990 or even later, which automatically limits the estimation period to some extent. One important aspect of this data set is the large fraction of quarterly data. Around one-third of variables are observed only in quarterly frequency. This is very atypical compared with similar studies for other countries that only use a very limited number of quarterly variables (GDP and possibly some quarterly surveys).

3. Forecasting Approaches

In light of the Swiss data situation and the choice of data set outlined in the previous section, the forecasting approaches under consideration need to be able to handle a large data set with different publication lags and frequencies. We evaluate three approaches: an *information combination* approach based on a DFM, a *model combination* approach pooling single-indicator models, and a specific-to-general *variable selection* procedure. In the following we briefly outline these three approaches. An online appendix (available at http://www.ijcb.org) provides more details on the models.

3.1 Information Combination Using a Dynamic Factor Model

The dynamic factor model approach models the co-movements of a panel of observed time series, x_t, as driven by a small number of latent factors, f_t, and idiosyncratic components, u_t, which can be represented as

[3]We checked if this intuition holds up by using two smaller indicator sets in our best-performing model (the dynamic factor model). See section 4.

[4]Indicators that are not available on a seasonally adjusted basis were calendar adjusted, seasonally adjusted, and outlier adjusted using the X-13ARIMA-SEATS procedure.

Table 1. Large-Scale Data Set for the Swiss Economy

Area	Soft Data		Hard Data		Prominent Examples
	M	Q	M	Q	
GDP				27	Total GDP, demand components, value added of sectors.
Labor Market	1	4	42	42	OASI statistics, unemployment statistics, job statistics, employment statistics, surveys, hours worked, wage index.
Consumption	4	13	6		Overnight stays of domestic visitors, retail sales, import of fuel, electricity consumption, new car registrations, consumer sentiment.
Investment			3	11	SwissMEM survey, imports of investment goods, industrial production of investment goods.
Foreign Trade			19	2	Trade statistics, overnight stays of foreign visitors.
International Activity	24	13	39	8	Several indicators covering Germany, euro area, United States, Japan, emerging Asia, and the CESifo world economic survey.
Financial Markets			64		Stock market indexes, exchange rates, commodity indexes, monetary aggregates, monetary conditions, interest rates, spreads.
Prices			12	3	Consumer prices, real estate prices, import prices, production prices, construction prices.
Construction Sector	15	3	6	21	Surveys, production, order books, cement deliveries.
Retail Trade Sector	6	5			Surveys.
Wholesale Trade Sector		14			Surveys.
Accommodation Sector		18	1		Overnight stays, surveys.
Manufacturing Sector	51	41	2	1	Industrial production, surveys, PMI, electricity production, number of working days.
Project Engineering Sector	8	7			Surveys.
Banking Sector	20	22	14		Credit statistics, surveys, balance sheet statistics, illiquidity index.
Insurance Sector	16	9			Surveys.
Other	7	2	1		Surveys.

$$x_t = \lambda f_t + u_t, \tag{1}$$

where f_t is an $r \times 1$ vector of common factors and λf_t is the common component. For the factors, we assume that they follow a finite-order Gaussian VAR,

$$f_t = \Phi_1 f_{t-1} + \cdots + \Phi_p f_{t-p} + v_t. \tag{2}$$

Both sets of equations constitute a state-space system. Doz, Giannone, and Reichlin (2012) show that the parameters and the latent factors can be estimated by maximum likelihood (ML) using the EM algorithm, which is a consistent estimator. Using the Kalman smoother in combination with the EM algorithm dominates principal components and two-step estimates.

Since the data set contains a large number of quarterly indicators, we use the modification suggested by Bańbura and Modugno (2014) allowing for mixed frequencies and arbitrary patterns of missing data. x_t is measured in monthly frequency (quarterly variables are treated as unobserved during the first two months of the quarter) and non-stationary variables are included as three-month growth rates.[5] Our baseline model uses two factors ($r = 2$) and one lag in the VAR of the factors ($p = 1$).[6,7]

---

[5]Giannone, Reichlin, and Small (2008), Angelini et al. (2011), and Bańbura and Rünstler (2011) also use three-month growth rates, whereas for example Bańbura and Modugno (2014) follow Mariano and Murasawa (2003) and stationarize trending monthly variables using month-on-month growth rates. We also evaluated the model using this transformation but found that the stationarization using three-month growth rates performs consistently better (see the online appendix).

[6]In choosing the number of factors, we try to obtain as parsimonious as possible a specification. While the inclusion of a second factor leads to a substantial increase in the correlation between GDP and its in-sample fit, a third and fourth factor do not lead to notable gains. As a robustness check, we estimated factor models using one and three factors and computed corresponding forecasts in the out-of-sample experiment. Although all findings of the baseline specification survived, two factors provide the best overall performance. The lag length in the VAR has no impact on the nowcasting performance and only very limited effects for the forecast. We therefore choose the most parsimonious specification, i.e., two factors and one lag.

[7]A different version of this model with four factors and estimated by the two-step procedure is proposed in Galli (2018) to obtain a business cycle index for the Swiss economy.

3.2 Model Combination Using MIDAS Regressions

This forecast approach pools the forecasts of several single-indicator models to obtain a final forecast. We first describe the single-indicator models and then present the combination method.

The single-indicator models relate h-steps-ahead GDP growth to the last p observations of the indicator using MIDAS and ARDL (autoregressive distributed lag) models,[8]

$$y^q_{t^q+h} = \begin{cases} \mu + \sum_{k=0}^{p^q} \beta_{k+1} x^q_{t^q+h-J-k} + u_{t^q+h} \\ \qquad\qquad \text{for quarterly indicators} \\ \mu + \sum_{k=0}^{p^m} \omega_k(\theta) x^m_{t^q+h,N^m-J-k} + u_{t^q+h} \\ \qquad\qquad \text{for monthly indicators} \\ \mu + \sum_{k=0}^{p^w} \omega_k(\theta) x^w_{t^q+h,N^w-J-k} + u_{t^q+h} \\ \qquad\qquad \text{for weekly indicators,} \end{cases} \tag{3}$$

where x denotes the indicator, which is available in quarterly (x^q), monthly (x^m) or weekly frequency (x^w). To allow for the most flexible dynamic specification, non-stationary variables are included as previous-period growth rates (stationary variables enter in levels).

For the quarterly frequency, the model is simply a distributed lag model of the indicator which is estimated by OLS without any restrictions. For the indicators available in monthly frequency, $N^m - J - k$ denotes the monthly lag of a specific indicator and $N^m = 3$ in this case (three monthly observations per quarter). J indicates the number of missing observations (in terms of months) for a specific quarter. k is measured in terms of months. For instance, when all monthly information for a given quarter is available, the forecasting horizon is $h = 1$, and six months (two quarters) are considered as lags, $k = 5$, $J = 0$, and $x^m_{t^q+1,3}, x^m_{t^q+1,2}, x^m_{t^q+1,1}, x^m_{t^q+1,0}, x^m_{t^q+1,-1}, x^m_{t^q+1,-2}$ are used as regressors in the model. This implies using six lags of monthly observations or a full set of information for quarter $t^q + 1$ and t^q.

[8]In a previous version of the paper, we allowed for lagged endogenous terms. However, this led to inferior forecasting results. Therefore, we opted for the simpler specification.

For weekly observations the same reasoning applies, although N^w is larger than N^m and time varying. Consequently, p^w is allowed to be larger than for quarterly and monthly indicators and J now measures the number of missings weeks of the respective quarter. The maximum lags for p^q, p^m, and p^w are set to match roughly one year of information; therefore, $p^q_{max} = 4$, $p^m_{max} = 12$, and $p^w_{max} = 48$.

For the higher-frequency indicators (monthly and weekly frequency) an Almon distributed lag model (Heinisch and Scheufele 2018) is applied. The function $\omega_k(\theta)$ is given by

$$\omega^i_k(\theta) = \omega^i_k(\theta_0, \theta_1, \dots \theta_q) = \theta_0 + \theta_1 k + \theta_2 k^2 + \cdots + \theta_q k^q. \qquad (4)$$

In this case the model can be estimated by restricted least squares. Even with a polynomial degree q that is substantially smaller than p, many functional forms can be well approximated. For each indicator, q together with p is selected by Bayesian information criterion (BIC) at each forecasting round (following Heinisch and Scheufele 2018).

To obtain the final forecast, we then pool the individual forecasts of each indicator: we take the average of all models with a BIC smaller than the one of the optimal AR model (GDP regressed on a constant and its own lags, where the number of lags is chosen by BIC).

3.3 Variable Selection Using a Specific-to-General Approach

A well-known alternative to information combination or model combination is variable selection. First, based on certain criteria, specific indicators from the large panel of indicators are selected. In a next step, they are then used in single equations. Typical examples are step-wise procedures that add (specific-to-general) or eliminate (general-to-specific) variables from a specification by means of information criteria or statistical tests (see, e.g., Ng 2013). Recent examples include Castle, Doornik, and Hendry (2011) and Chudik, Kapetanios, and Pesaran (2016).

We use a modified version of the specific-to-general approach suggested by Herwartz (2011a, 2011b): when selecting the variables we correct the significance levels for multiple testing similar to Chudik, Kapetanios, and Pesaran (2016). Moreover, we apply blocking (see, e.g., McCracken, Owyang, and Sekhposyan 2015) and realignment to allow for mixed frequencies and different publication lags. The algorithm works as follows:

(i) The blocking and realignment yields a balanced quarterly panel of candidate indicators, W (non-stationary variables enter in terms of period-wise growth rates). The set of selected indicators, W^{sg}, is initialized with only a constant.

(ii) Project GDP growth onto W^{sg} and retain the estimated residuals $\hat{u}$.

(iii) Separately regress $\hat{u}$ on each candidate indicator in W. For each regression, calculate the Lagrange multiplier (LM) statistic.

(iv) If the highest LM statistic is above the critical value computed according to the Benjamini and Hochberg (1995) procedure that controls for multiple testing, move the corresponding indicator from W to W^{sg} and return to step (ii). Otherwise proceed to step (v).

(v) Use the last iteration's step (ii) projection to generate the final GDP forecast.

3.4 Benchmark Models

As benchmark models, we consider (i) optimal AR models and (ii) popular leading indicators. The AR model has a maximum of four lags, with the lag length being optimized at each forecast round using the BIC. For the leading-indicator models, we consider the two most prominent leading indicators for the Swiss economy: the PMI in the manufacturing sector and the KOF economic barometer. Both indicators are released on a monthly basis, receive some attention in the public, and are promising in terms of forecasting performance (see, e.g., Maurer and Zeller 2009, Lahiri and Monokroussos 2013, and Abberger et al. 2014). For the two leading indicators, we use the same MIDAS framework as outlined in section 3.2, where the specification is selected using BIC. Both the AR and the leading-indicator models are optimized separately for each forecast horizon.

4. Model and Forecast Evaluation

To evaluate the models, we produce a one- and two-quarter-ahead forecast using a recursive pseudo-real-time setup.[9] The different models are compared in terms of root mean squared error (RMSE). The target variable is the annualized quarter-on-quarter growth of Swiss GDP (real, calendar, and seasonally adjusted).

Two specific dates are selected to time the forecast production. The first, "early-quarter," information set is given by indicator information available on the 11th of March/June/September/December. At this date, surveys are usually available for one month of the nowcast quarter, hard data for one month or not at all, and financial data for two months.[10] For this information set, we use an underlying data vintage available on September 11, 2015.

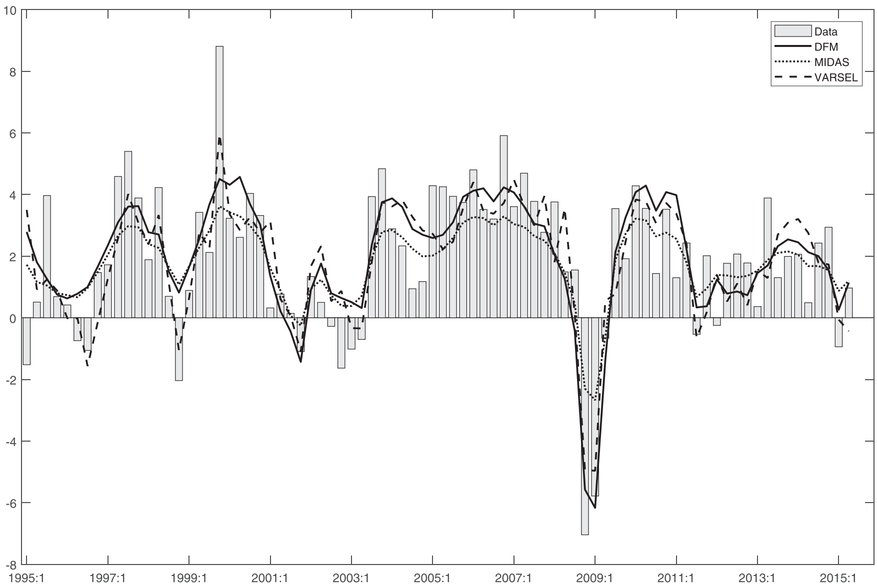
The second, "late-quarter," information set is given by data available on the 6th of May/August/November/February. At this date, surveys are usually available for all three months of the nowcast quarter, hard data for two or three months, and financial data for all three months. Theoretically, this additional information should improve the forecasts. For this information set, we use an underlying data vintage available on November 6, 2015.

The evaluation is conducted over the period 2005:Q1–2015:Q2. It is based on GDP data that came with the State Secretariat for Economic Affairs' official GDP release for the second quarter of 2015. In what follows, the forecast horizon will always be defined relative to the GDP data situation. This means that if, e.g., GDP is available up to Q4, the one-step-ahead forecast (i.e., the nowcast, defined as $h = 1$) will—for both information sets (March 11 and May 6)—target Q1 GDP growth and the two-step-ahead forecast ($h = 2$) Q2 GDP growth.

[9]Note that such a pseudo-real-time setup does not take into account the fact that, in reality, indicators may be subject to revisions, including revisions stemming from the seasonal adjustment procedure.

[10]Kaufman and Scheufele (2017) investigate the informational content of the surveys by the KOF Swiss Economic Institute—one of the main sources for business tendency data on Switzerland. They find that these surveys are helpful for explaining inflation, employment growth, and the output gap, but less so for GDP growth.

Figure 1. In-Sample Fit of the Three Approaches under Investigation



Note: This graph shows the in-sample explanatory power of three models along with quarterly annualized GDP growth.

4.1 In-Sample Fit

Figure 1 plots the in-sample fit of the three approaches under investigation against realized GDP growth.[11] Several observations are in order. First, the models' general interpretations of the business cycle are similar: all models identify phases of weakness around 1996, 1999, 2002, and 2009. Also, the models agree that there were significant decelerations during 2011 (strong CHF appreciation and a global slowdown) and at the beginning of 2015 (strong CHF appreciation following the end of the EUR–CHF exchange rate floor). Second, the

[11]While the computation of the DFM in-sample fit is straightforward, it is less obvious how to compute it based on MIDAS and VARSEL, because these are direct forecasting approaches. For MIDAS and VARSEL we therefore estimate the indicator models using a balanced data set. This implies that for a given quarter of GDP we take all three months of indicator information as given.

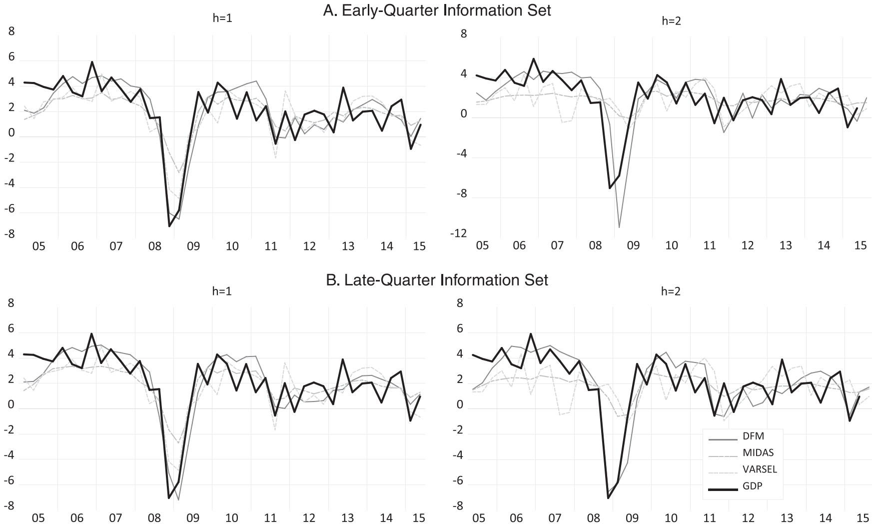
models also agree on a number of quarters, where GDP appears to be driven by idiosyncrasies that are not mirrored in the broad data set from which the model dynamics are constructed. Examples are the strong GDP growth rates of 1995:Q1, 1999:Q3, and 2014:Q4 and the weak GDP growth in 2004:Q3 and 2004:Q4, 2010:Q3, 2011:Q1, and 2014:Q2. This illustrates the usefulness of a large data set for measuring the underlying business cycle dynamics in a broader sense than what is captured by GDP. Third, the smoothness of the in-sample fit varies considerably, with the model combination approach (MIDAS) being particularly smooth (standard deviation of 1.3). The other two models—the dynamic factor model (DFM) and the variable selection approach (VARSEL)—produce a more volatile fit, with standard deviations of approximately 1.9. For comparison, the standard deviation of GDP is 2.4 percentage points. Fourth, the in-sample fit varies only slightly across models. MIDAS has an in-sample R^2 of 0.71. The two other models' R^2 are 0.63 (VARSEL) and 0.65 (DFM). Whether these good in-sample fits translate into a good forecasting performance will be addressed in the next section.

4.2 Out-of-Sample Evaluation Results

Figure 2 plots the forecasts of the three approaches under investigation over time. The main features of the models are directly visible in these graphs. For the nowcast ($h = 1$), all of the investigated models more or less follow the tendency of GDP growth. Clearly, there are periods where the connection is stronger (e.g., before, during, and after the financial crisis) and periods where GDP growth deviates more persistently from the predictions of the models (in 2005 and 2013). This broadly corresponds to the in-sample results. Generally, forecasts of the different models are highly correlated (particularly for $h = 1$). Additionally, forecasts using the early-quarter information set are not much different from the late-quarter information set. As expected, the forecasts are more accurate for the nowcast ($h = 1$) than for the following quarter ($h = 2$).

Looking at the different models, we see that the DFM captures the financial crisis period 2008/09 very well. The model combination strategy (MIDAS) is less successful during the crisis than the DFM, but at least gets the tendency right. It generally delivers less volatile forecasts and is therefore less flexible in situations of rapid

Figure 2. Model Forecasts and Realizations



Notes: The graphs show the recursively computed annualized GDP forecasts of different models for two forecasting horizons (h = 1 and h = 2) and realized GDP growth. Panel A is based on the early-quarter information set and panel B is based on the late-quarter information set.

change. The variable selection approach also performs quite well for the nowcast ($h = 1$) during the crisis period. However, this forecasting method is very volatile for $h = 2$ and seems to be less correlated with GDP growth than most of the other forecasting models.

4.2.1 Average Performance

Table 2 reports the relative root mean squared errors of the models against the AR benchmark. Although the forecast errors for the post-crisis sample—which covers 2010:Q1–2015:Q2—are, on average, smaller (see, e.g., the AR benchmark) in absolute terms, the relative performance of the models under investigation has deteriorated somewhat (relative to the AR benchmark). A similar result is well documented for the pre-crisis period relative to the crisis period (Drechsel and Scheufele 2012; Kuzin, Marcellino, and Schumacher 2013).

Table 2. Relative RMSEs of Different Forecasting Methods

	Total Sample 2005:Q1–2015:Q2		Post-Crisis Sample 2010:Q1–2015:Q2	
	$h = 1$	$h = 2$	$h = 1$	$h = 2$
AR Benchmark	2.1689	2.5081	1.6334	1.2686
A. Early-Quarter Information				
a. Information Combination				
DFM	0.666*	0.806*	0.909	1.221
b. Model Combination				
MIDAS	0.768**	0.870**	0.760***	0.927
c. Variable Selection				
VARSEL	0.726*	1.056	0.936	1.470
d. Leading Indicator Models				
PMI	0.829	0.871*	1.001	1.055
KOF Barometer	0.850	0.844	0.923	0.958
e. Pooling of Different Procedures				
COMBO1	0.654**	0.701**	0.809**	0.999
COMBO2	0.641**	0.782*	0.785**	1.112
COMBO3	0.630*	0.792*	0.838*	1.261
COMBO4	0.700**	0.942	0.776**	1.149
B. Late-Quarter Information				
a. Information Combination				
DFM	0.656*	0.588*	0.833*	1.114
b. Model Combination				
MIDAS	0.719**	0.818**	0.704***	0.922
c. Variable Selection				
VARSEL	0.726*	1.059	0.937	1.486
d. Leading Indicator Models				
PMI	0.733*	0.845	0.899	1.359
KOF Barometer	0.938	0.834	1.029	1.057

(*continued*)

Table 2. (Continued)

	Total Sample 2005:Q1–2015:Q2		**Post-Crisis Sample 2010:Q1–2015:Q2**	
	$h = 1$	$h = 2$	$h = 1$	$h = 2$
e. Pooling of Different Procedures				
COMBO1	0.619**	0.588**	0.741***	0.930
COMBO2	0.615**	0.711**	0.741***	1.039
COMBO3	0.617**	0.695*	0.797**	1.143
COMBO4	0.676**	0.918	0.753***	1.153

Notes: The table shows relative root mean squared errors (RMSEs) for $h = 1$ (nowcast) and $h = 2$ (one additional quarter ahead) using two different states of information. Besides the benchmark AR model, all numbers are defined relative to the benchmark. Pooling of different procedures includes the following models using equal weights. COMBO1: DFM + MIDAS, COMBO2: DFM + MIDAS + VARSEL, COMBO3: DFM + VARSEL, COMBO4: MIDAS + VARSEL. ***, **, and * indicate whether a model's predictive ability (using the DM test) is significantly better than the benchmark (at the 1 percent, 5 percent, and 10 percent level, respectively).

Overall, the DFM performs best in the total evaluation sample. The forecast gains against the AR benchmark models are more than 30 percent. Compared with the leading-indicator models, the DFM offers improvements between 10 and 20 percent. In the post-crisis sample, the gains relative to the AR decline to 10–20 percent. The improvements in forecasting accuracy from the early-quarter to the late-quarter information set are, on average, small. The gains from the early to the late information set are only 1 percent and 5 percent, for the DFM and MIDAS, respectively. For $h = 2$, the gains are slightly larger.

Among the approaches under investigation, the forecast gains of the DFM against MIDAS and VARSEL are around 10 percent for $h = 1$ (see table 3 for pairwise model comparisons). None of the three models is able to systematically outperform the other two approaches. When considering the pooling of different methods (COMBO1 and COMBO2), we find that those are on average often better than the individual procedures and do significantly outperform the variable selection procedure (VARSEL). For $h = 2$, VARSEL performs very poorly compared with the other two produres and the pooling of methods.

Table 3. Pairwise Forecast Comparisons

Bench	$h = 1$				$h = 2$			
	DFM	**MIDAS**	**VARSEL**	**COMBO1**	**DFM**	**MIDAS**	**VARSEL**	**COMBO1**
A. Early-Quarter Information								
MIDAS	1.15				1.08			
VARSEL	1.09	0.94			1.31***	1.21***		
COMBO1	0.98	0.85	0.90		0.87	0.81	0.66***	
COMBO2	0.96	0.83	0.88**	0.98	0.97	0.90	0.74***	1.11***
Best Model (in % of time)	35.7	45.2	19.1	4.8	41.5	22.0	17.1	19.5
B. Late-Quarter Information								
MIDAS	1.10				1.39			
VARSEL	1.11	1.01			1.80**	1.29***		
COMBO1	0.94	0.86	0.85*		1.00	0.72	0.56**	
COMBO2	0.94	0.86	0.85***	0.99	1.21	0.87	0.67***	1.21**
Best Model (in % of time)	31.0	38.1	14.3	9.5	36.6	31.7	14.6	14.6

Notes: Results of pairwise comparisons are displayed using relative RMSEs and the test for equal predictive ability of DM. The benchmark is given by the top line (and is used as the denominator of the relative RMSE). ***, **, and * indicate whether the DM test is significant at the 1 percent, 5 percent, and 10 percent level, respectively. Additionally, we show the percentage of times when models provide the best forecast.

Interestingly, when the focus is not on the average performance in terms of RMSEs, but on picking the best model as often as possible (best model in percentage of time; see table 3), the ranking of models changes. For $h = 1$, MIDAS is most often the model that is closest to the realizations, followed by the DFM. For $h = 2$, the order is turned around, with DFM slightly more successful. Pooling of different methods is not successful in this case.

The forecasting performance of model combination based on MIDAS models is slightly less accurate for the total sample compared with the DFM. The forecasting gains in terms of RMSEs are approximately 23 percent (early-quarter information) and 28 percent (late-quarter information). This is broadly in line with previous research that compares similar methods (Bańbura et al. 2013; Kuzin, Marcellino, and Schumacher 2013; Foroni and Marcellino 2014). However, for the post-crisis sample and $h = 1$, model combination based on MIDAS models offers the largest and most significant improvement relative to the benchmark model.

Interestingly, the specific-to-general variable selection approach (VARSEL) performs reasonably well for the nowcast period (and is able to outperform the benchmark for the total sample). For one quarter ahead ($h = 2$), however, this approach is not very reliable, and for the post-crisis sample there are no significant improvements relative to the benchmark.

It is also instructive to assess how three approaches under investigation compare with single-leading-indicator models. For the total sample, the performance of the PMI and KOF economic barometer is less accurate than the DFM and slightly inferior to the model combination and variable selection approaches. Only the PMI offers some significant improvements relative to the benchmark. The performances of the two leading-indicator models have clearly deteriorated in the post-crisis period, and the models contain only little information. Most interestingly, the KOF economic barometer—which is itself calculated from a large-scale factor model—does not outperform the much simpler PMI.

4.2.2 Varying the Size of the Data Set

The fact that we use a very large data set seems to be controversial. Boivin and Ng (2006) for the United States and Caggiano,

Table 4. Relative RMSEs of DFMs Using a Smaller Data Set

	Early-Quarter Information Set		Late-Quarter Information Set	
	$h = 1$	$h = 2$	$h = 1$	$h = 2$
Baseline	0.666	0.806	0.656	0.588
113 Indicators	0.719	0.809	0.682	0.722
20 Indicators	0.733	1.037	0.705	0.877

Notes: The table shows relative root mean squared errors (RMSEs) against the AR benchmark for $h = 1$ (nowcast) and $h = 2$ (one additional quarter ahead) using two different states of information. "Baseline" corresponds to the model considered above; "113 Indicators" and "20 Indicators" use smaller data sets.

Kapetanios, and Labhard (2011) for some euro-area countries show that including a very large set of variables in the factor model might lead to inferior forecasting results. Moreover, Bańbura, Giannone, and Reichlin (2011) and Bańbura and Modugno (2014) find that including disaggregate information does not lead to an improved forecasting performance for euro-area GDP forecasts.

Therefore, we compare the best-performing model (the DFM estimated with the EM algorithm) using the large data set with DFMs using a medium-sized and a smaller data set. The smaller data set contains twenty variables that were selected using expert judgment, whereas the medium-sized data set contains 113 indicators that were selected as follows: we started with our large data set and went through each of the indicator categories outlined in table 1 of the paper. For categories with more than ten monthly indicators, we deleted the quarterly indicators and some details of the monthly indicators. For categories with less than ten monthly indicators, we deleted details of the quarterly indicators. The goal was the get about five to ten indicators per category.

Table 4 reports the results. The empirical performance of the medium-sized data set is lower than that of the large data set, although the differences are not very large and also the medium-sized data set still clearly beats all benchmarks. The smaller data set performs clearly worse than the large data set and also worse than the medium-sized data set.

The fact that more indicators appear to be better for Switzerland is most likely driven by the relatively unsatisfactory data situation in Switzerland, which we describe in section 2: many indicators that economists routinely use to track economic developments in the United States (or also in the euro zone) are simply not available for Switzerland or, if they are, they are only available on a quarterly frequency, with a long publication lag and/or a short history. Therefore, it seems to be a good strategy to use whatever is available and go with a large data set.

4.2.3 Time-Varying Forecasting Performance

Figure 3 gives a more complete picture of the time-varying performance of the models. First, it confirms that the good performance of the factor model mainly comes from the crisis period 2008/09. Second, it shows that the performance of the three approaches under investigation relative to the AR benchmark has deteriorated in 2010–13. Third, most recently, the models' performance gained again, most likely due to the exchange rate shock in 2015. Interestingly, the performance of the model combination approach based on MIDAS models remains very stable over time and is less sensitive.

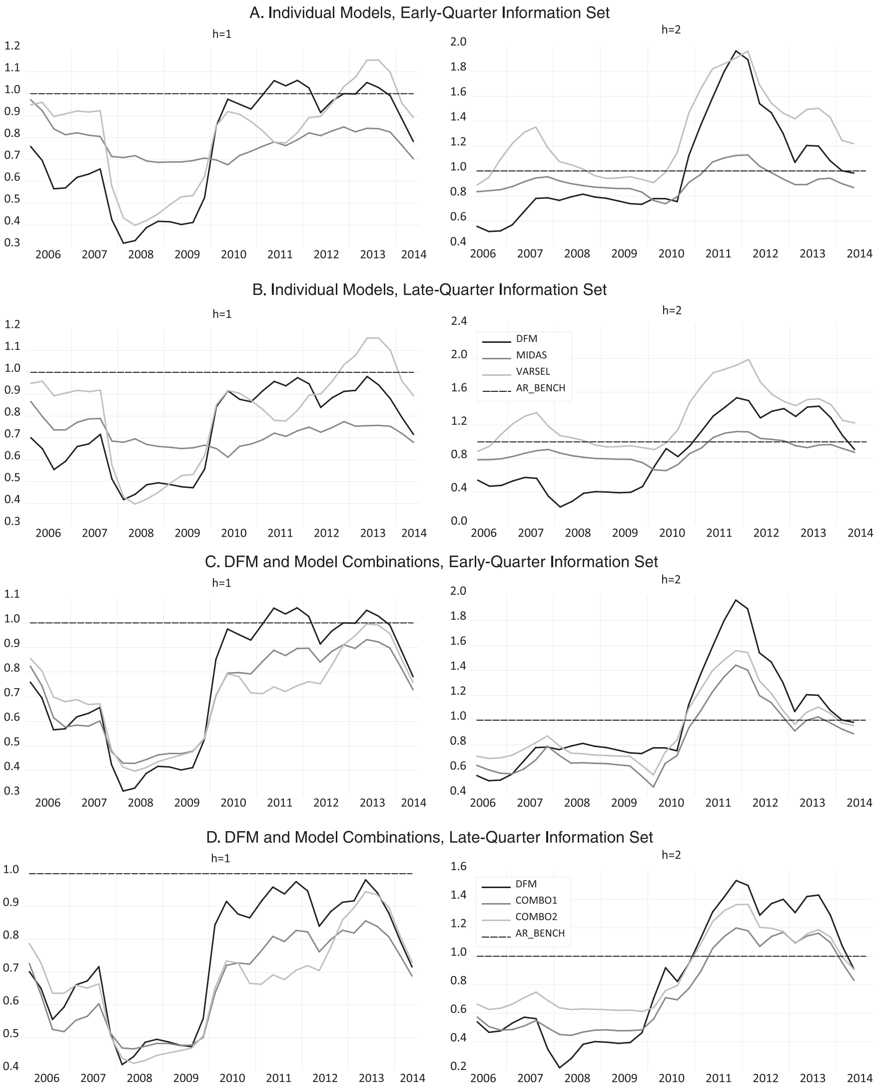
4.2.4 Pooling Different Forecasting Models

Table 2 and figure 3 show that one may obtain additional gains in average performance and in terms of reliability (significance) when pooling different forecast methods. For the total sample and the nowcast period, just pooling the DFM and MIDAS (COMBO1) results in a very good performance. In some cases, pooling all three methods (COMBO2) improves the results even further. Generally, the different poolings perform very similarly, and it is difficult to identify a clear winner. Pooling of mixed-frequency model forecasts is stable and reliable in terms of forecast accuracy. Additionally, figure 3 shows that, since 2010, the pooling of different procedures is superior to the single DFM approach.

4.2.5 Forecasting the Tendency of Output Growth

When looking at other measures of forecasting performance, our general results hold. However, the details may be slightly different. By

Figure 3. Rolling Relative RMSEs (Two-Year Windows)



Notes: The graphs show eight quarters centered moving average of a model's RMSE relative to the AR benchmark model. Values smaller than one indicate that the local forecasting performance is better than the benchmark model.

looking at whether the models are able to accurately capture the general tendency of output growth, we go from a continuous measure (in our case, the RMSE), to a nominal measure. In practice, the forecaster may not be judged by the specific distance of the forecast to a target, but by whether he got the general tendency right. By using a trichotomous measure, we take into account three different phases: low, moderate, and strong growth states. We can then investigate how our forecasts match those categories when compared with the realizations. Our definition of the three different phases is motivated by two different considerations: first, by the empirical distribution of GDP growth since the 1990s, and second, by policy considerations. A policymaker may be interested in whether the economy expands robustly, which we define by growth that is markedly greater than the long-term average of 1.6 percent—so we opt for the 2 percent threshold. Additionally, a policymaker might become worried when growth is only marginally positive or even negative. This motivates our second threshold of 0.5 percent.

Table 5 shows the main results for the nowcast period. Generally, the main findings are compatible with previous results. First, the differences among the different approaches tend to be small. Second, all models under investigation are informative. Thus, the null hypothesis of independence between forecasts and realizations can be rejected in all cases. On average, the predictions match the corresponding realized categories in approximately 52–64 percent of the cases. Third, there seems to be only a very small improvement in terms of accuracy from the early to the late information set. Fourth, model averaging does not improve the performance. Overall, the DFM provides the best results in terms of predicting the business cycle phase (correlation coefficient).

5. Conclusions

Monitoring economic developments in real time is one of the most important but also most challenging tasks that the applied economist working on Switzerland faces. In this paper, we set up a large database containing hundreds of potentially relevant variables. We then considered different approaches to condensing the information of the data set into a GDP forecast. The traditional approaches often used in practice select one or a few indicators based on expert

Table 5. Contingency Tables for Business Cycle Phase Predictions

A. Early-Quarter Information Set						
	a. DFM			**b. MIDAS**		
	Realization			**Realization**		
	(−∞, 0.5]	**(0.5, 2]**	**(2, +∞)**	**(−∞, 0.5]**	**(0.5, 2]**	**(2, +∞)**
Prediction						
(−∞, 0.5]	11.9	4.8	2.4	7.1	0.0	2.4
(0.5, 2]	4.8	4.8	14.3	9.5	11.9	16.7
(2, +∞)	2.4	14.3	40.5	2.4	11.9	38.1
% of Correct Predictions			57.14			57.14
Correlation Coefficient			0.59			0.55
Pearson χ^2			14.40			12.80
p Value			0.006			0.012
	d. VARSEL			**e. COMBO1**		
	Realization			**Realization**		
	(−∞, 0.5]	**(0.5, 2]**	**(2, +∞)**	**(−∞, 0.5]**	**(0.5, 2]**	**(2, +∞)**
Prediction						
(−∞, 0.5]	11.9	7.1	0.0	11.9	2.4	2.4
(0.5, 2]	4.8	4.8	16.7	4.8	7.1	16.7
(2, +∞)	2.4	11.9	40.5	2.4	14.3	38.1
% of Correct Predictions			57.14			57.14
Correlation Coefficient			0.64			0.61
Pearson χ^2			17.15			15.93
p Value			0.002			0.003
B. Late-Quarter Information Set						
	a. DFM			**b. MIDAS**		
	Realization			**Realization**		
	(−∞, 0.5]	**(0.5, 2]**	**(2, +∞)**	**(−∞, 0.5]**	**(0.5, 2]**	**(2, +∞)**
Prediction						
(−∞, 0.5]	11.9	4.8	2.4	7.1	2.4	2.4
(0.5, 2]	4.8	7.1	9.5	11.9	7.1	16.7
(2, +∞)	2.4	11.9	45.2	0.0	14.3	38.1
% of Correct Predictions			64.29			52.38
Correlation Coefficient			0.65			0.55
Pearson χ^2			17.56			12.86
p Value			0.002			0.012

(continued)

Table 5. (Continued)

B. Late-Quarter Information Set						
	d. VARSEL			**e. COMBO1**		
	Realization			**Realization**		
	$(-\infty$, 0.5]	**(0.5, 2]**	**(2, $+\infty$)**	**$(-\infty$, 0.5]**	**(0.5, 2]**	**(2, $+\infty$)**
Prediction						
$(-\infty$, 0.5]	11.9	7.1	0.0	9.5	2.4	2.4
(0.5, 2]	4.8	4.8	16.7	7.1	7.1	14.3
(2, $+\infty$)	2.4	11.9	40.5	2.4	14.3	40.5
% of Correct Predictions			57.14			57.14
Correlation Coefficient			0.64			0.55
Pearson χ^2			17.15			12.92
p Value			0.002			0.012

Notes: Results of pairwise comparisons about three different business cycle phases. The numbers reflect frequencies (in percentage points) how predictions match with realizations. 1: low/negative growth (below 0.5 percent), 2: moderate growth (0.5–2 percent), and 3: high growth (above 2 percent). Additionally, statistics about the degree of association are displayed (as well as a test of independence).

knowledge and derive the forecast using OLS regressions or a small-scale dynamic factor model. Alternatively, one may use all indicators without an expert's pre-selection. We presented and compared three approaches to doing so.

The first approach, factor-based *information combination*, extracts a small number of common factors from the database and forms GDP forecasts based on these factors. It is implemented by a Kalman-filter-based DFM approach which is estimated by the EM algorithm. This method provides very good results for nowcasting GDP and beats relevant benchmarks such as univariate time-series models, prominent leading-indicator models, and a small-scale factor model.

The second approach, *model combination*, performs estimations based on MIDAS equations for each indicator and then combines the resulting forecasts to form a final GDP forecast. This makes communicating a particular indicator's contribution to the forecast very easy. Also, the variation in the distribution of the forecasts over time can be used as a measure of uncertainty. Moreover, it is quite easy to implement. This model performs slightly worse than the dynamic factor model within the total evaluation sample, mainly because it is not fully able to catch the large drop in output during the financial crisis. For the post-crisis sample, however, its forecasting performance is even slightly better than the factor model.

Our third approach to extract relevant information from the large data set is *variable selection*. A specific-to-general approach that can handle very large data sets and mixed frequencies is used for this purpose. Although this approach delivers forecasts that are less accurate than those of the information combination approach and the model combination approach, in certain cases, it still offers some improvements against the benchmark.

Additionally, we experimented with pooling the three approaches' forecasts. Forecast pooling of two or three methods delivers very reliable forecasts. Overall, this suggests that it is best to employ several short-term forecast methods on a large data set and to pool the results of the most-promising methods. This is much better than relying on one method that uses only one indicator, and it is particularly beneficial after the financial crisis. Given this finding, we strongly recommend using a large data set and several forecasting approaches when monitoring the Swiss economy.

References

Abberger, K., M. Graff, B. Siliverstovs, and J.-E. Sturm. 2014. "The KOF Economic Barometer, Version 2014: A Composite Leading Indicator for the Swiss Business Cycle." KOF Working Paper No. 353, KOF Swiss Economic Institute.

Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler. 2011. "Short-term Forecasts of Euro Area GDP Growth." *Econometrics Journal* 14 (1): C25–C44.

Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin. 2013. "Now-casting and the Real-time Data Flow." In *Handbook of Economic Forecasting*, Vol. 2A, ed. G. Elliot and A. Timmermann, 195–237 (chapter 4). Elsevier.

Bańbura, M., D. Giannone, and L. Reichlin. 2011. "Nowcasting." In *The Oxford Handbook of Economic Forecasting*, ed. M. P.

Clements and D. F. Hendry, 193–224 (chapter 7). Oxford University Press.

Bańbura, M., and M. Modugno. 2014. “Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data.” *Journal of Applied Econometrics* 29 (1): 133–60.

Bańbura, M., and G. Rünstler. 2011. “A Look into the Factor Model Black Box: Publication Lags and the Role of Hard and Soft Data in Forecasting GDP.” *International Journal of Forecasting* 27 (2): 333–46.

Benjamini, Y., and Y. Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society, Series B* 57 (1): 289–300.

Boivin, J., and S. Ng. 2006. “Are More Data Always Better for Factor Analysis?” *Journal of Econometrics* 132 (1): 169–94.

Caggiano, G., G. Kapetanios, and V. Labhard. 2011. “Are More Data Always Better for Factor Analysis? Results for the Euro Area, the Six Largest Euro Area Countries and the UK.” *Journal of Forecasting* 30 (8): 736–52.

Castle, J. L., J. A. Doornik, and D. F. Hendry. 2011. “Evaluating Automatic Model Selection.” *Journal of Time Series Econometrics* 3 (1): 1–33.

Chudik, A., G. Kapetanios, and M. H. Pesaran. 2016. “Big Data Analytics: A New Perspective.” Working Paper.

Doz, C., D. Giannone, and L. Reichlin. 2012. “A Quasi–Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models.” *Review of Economics and Statistics* 94 (4): 1014–24.

Drechsel, K., and R. Scheufele. 2012. “The Performance of Short-term Forecasts of the German Economy before and during the 2008/2009 Recession.” *International Journal of Forecasting* 28 (2): 428–45.

Evans, M. D. D. 2005. “Where Are We Now? Real-Time Estimates of the Macroeconomy.” *International Journal of Central Banking* 1 (2): 127–75.

Foroni, C., and M. Marcellino. 2014. “A Comparison of Mixed Frequency Approaches for Nowcasting Euro Area Macroeconomic Aggregates.” *International Journal of Forecasting* 30 (3): 554–68.

Galli, A. 2018. “Which Indicators Matter? Analyzing the Swiss Business Cycle Using a Large-Scale Mixed-Frequency Dynamic Factor Model.” *Journal of Business Cycle Research* 14 (2): 179–218.

Giannone, D., L. Reichlin, and D. Small. 2008. “Nowcasting: The Real-time Informational Content of Macroeconomic Data Releases.” *Journal of Monetary Economics* 55 (4): 665–76.

Heinisch, K., and R. Scheufele. 2018. “Bottom-up or Direct? Forecasting German GDP in a Data-rich Environment.” *Empirical Economics* 54 (2): 705–45.

Herwartz, H. 2011a. “Forecast Accuracy and Uncertainty in Applied Econometrics: A Recommendation of Specific-to-General Predictor Selection.” *Empirical Economics* 41 (2): 487–510.

———. 2011b. “Specific-to-General Predictor Selection in Approximate Autoregressions—Monte Carlo Evidence and a Large-Scale Performance Assessment with Real Data.” *AStA Advances in Statistical Analysis* 95 (2): 147–68.

Kaufmann, D., and R. Scheufele. 2017. “Business Tendency Surveys and Macroeconomic Fluctuations.” *International Journal on Forecasting* 33 (4): 878–93.

Kuzin, V., M. Marcellino, and C. Schumacher. 2011. “MIDAS vs. Mixed-Frequency VAR: Nowcasting GDP in the Euro Area.” *International Journal of Forecasting* 27 (2): 529–42.

———. 2013. “Pooling versus Model Selection for Nowcasting GDP with Many Predictors: Empirical Evidence for Six Industrialized Countries.” *Journal of Applied Econometrics* 28 (3): 392–411.

Lahiri, K., and G. Monokroussos. 2013. “Nowcasting US GDP: The Role of ISM Business Surveys.” *International Journal of Forecasting* 29 (4): 644–58.

Mariano, R. S., and Y. Murasawa. 2003. “A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series.” *Journal of Applied Econometrics* 18 (4): 427–43.

Maurer, C., and M. Zeller. 2009. “PMI — aktuelles Konjunkturbarometer der Schweiz.” *Die Volkswirtschaft* 3-2009: 51–52.

McCracken, M. W., M. T. Owyang, and T. Sekhposyan. 2015. “Real-time Forecasting with a Large, Mixed Frequency, Bayesian VAR.” Working Paper No. 2015-030A, Federal Reserve Bank of St. Louis.

Ng, S. 2013. "Variable Selection in Predictive Regressions." In *Handbook of Economic Forecasting*, Vol. 2, Part B, ed. G. Elliott and A. Timmermann, 753–89 (chapter 14). Elsevier.

Stock, J. H., and M. W. Watson. 2002. "Macreconomic Forecasting Using Diffusion Indexes." *Journal of Business and Economic Statistics* 20 (2): 147–62.

———. 2011. "Dynamic Factor Models." In *The Oxford Handbook of Economic Forecasting*, ed. M. P. Clements and D. F. Hendry, 35–59 (chapter 2). Oxford University Press.