Online Appendix to "Mixed-Frequency Models for Tracking Short-Term Economic Developments in Switzerland"

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

Introduction

This online appendix provides a more detailed exposition of the models under consideration in "Mixed-Frequency Models for Tracking Short-Term Economic Developments in Switzerland." It also provides some additional results that were not reported in the main text.

Information Combination Using a Dynamic Factor Model

As outlined in the main text, we use the EM algorithm as suggested by Doz, Giannone, and Reichlin (2012) and extended by Bańbura and Modugno (2014) to allow for mixed frequencies and arbitrary patterns of missing data. In the following, we describe the model and our particular specification choices in somewhat more detail, explain how we initialize the EM algorithm, and briefly sketch the empirical performance of some alternative model specifications.

Some Details on the Model and Specification

We define our data vector at a monthly frequency as

$$z_t = \begin{bmatrix} x_{q,t}^{Q} \\ x_{m,t}^{M} \end{bmatrix}, \tag{1}$$

where the superscripts Q and M denote whether the variable is observed in monthly or quarterly frequency and the subscripts q

and m denote at which frequency the variable is actually computed. For GDP, e.g., $X^Q_{i=gdp,q,t}$ denotes the observed quarterly value for GDP in period t, and $X^Q_{i=gdp,m,t}$ denotes the (latent) monthly value for GDP in period t. The vector $x^Q_{q,t}$ collects the n^Q quarterly indicators (observed only in the third month of the quarter) including GDP, whereas the vector $x^M_{m,t}$ stacks the n^M monthly indicators. The total number of indicators is $n = n^Q + n^M$.

We stationarize trending quarterly variables using three-month growth rates following Rünstler et al. (2009). Given this choice, we can write the vector of observed indicators z_t as

$$z_t = (G_0 + G_1 L + G_2 L^2) \begin{bmatrix} x^Q_{m,t} \\ x^M_{m,t} \end{bmatrix}, \tag{2}$$

where $x^Q_{m,t}$ is unobserved and

$$G_0 = \begin{bmatrix} \frac{1}{3} I_{n^Q} & \mathbf{O} \\ \mathbf{O} & I_{n^M} \end{bmatrix}, G_1 = \begin{bmatrix} \frac{1}{3} I_{n^Q} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}, G_2 = \begin{bmatrix} \frac{1}{3} I_{n^Q} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}. \tag{3}$$

Defining $\mathbf{G}(L) = (G_0 + G_1 L + G_2 L^2)$, we can write the factor model as $z_t = \mathbf{G}(L)(\Lambda f_t + u_t)$. In state-space form we obtain the following measurement equation:

$$z_t = H_t \underbrace{\begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \end{bmatrix}}_{\substack{\xi_t \\ 3r \times 1}} + \underbrace{\begin{bmatrix} \tilde{u}^Q_t \\ u^M_t \end{bmatrix}}_{\epsilon_t}, \tag{4}$$

with $\epsilon_t \sim N(\mathbf{0}_n, R_t)$. $\tilde{u}^Q_t = G(L) u^Q_t$ is assumed to be white noise.[1] Note that in z_t, we replace missing values with zeros. Therefore, H_t and R_t are modifications of the matrices

[1]Conceptionally, u^Q_t would actually follow a moving-average process. However, as outlined in Angelini, Bańbura, and Rünstler (2010), this does not affect the consistency of the estimates. Furthermore, without this assumption, the EM algorithm would not be applicable to factor models with mixed frequencies, as mentioned in Mariano and Murasawa (2010).

$$H = \begin{bmatrix} G_0\Lambda & G_1\Lambda & G_2\Lambda \end{bmatrix} \text{ and } R = \begin{bmatrix} \frac{3}{9}\Sigma_u^Q & \mathbf{O} \\ \mathbf{O} & \Sigma_u^M \end{bmatrix}, \tag{5}$$

with their elements adjusted according to the data situation of period t. That is, if an indicator i is unobserved at time period t, the i-th line of H is replaced with zeros to obtain H_t and the i-th entry of the diagonal of R is replaced with 0 to obtain R_t. Note that this procedure applies to quarterly indicators that are unobserved because t is the first or second month of a quarter as well as to unobserved monthly indicators. More details on how the Kalman filter is able to address missing observations can be found in Durbin and Koopman (2012).

The transition equation representing the VAR of the factors is given by

$$\begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \end{bmatrix} = \begin{bmatrix} \Phi_{r\times r} & \mathbf{O}_{r\times 2r} \\ I_{2r} & \mathbf{O}_{2r\times r} \end{bmatrix} \begin{bmatrix} f_{t-1} \\ f_{t-2} \\ f_{t-3} \end{bmatrix} + e_t, \tag{6}$$

where

$$e_t = \begin{bmatrix} v_t \\ \mathbf{0}_{2r} \end{bmatrix} \sim N\left(\mathbf{0}_{3r}, Q\right), \text{ and } Q = \begin{bmatrix} I_r & \mathbf{O}_{r\times 2r} \\ \mathbf{O}_{2r\times r} & \mathbf{O}_{2r\times 2r} \end{bmatrix}. \tag{7}$$

Obtaining the Starting Values for the EM Algorithm

The model is estimated using the EM algorithm as outlined in Bańbura and Modugno (2014). The starting values for the algorithm are obtained as follows:

(i) Perform a principal component analysis with monthly variables only to obtain the first r principal components f_t^0.

(ii) Regress all variables on the first r principal components to get an initial estimate for Λ:
- For quarterly variables: $x_{i,q,t}^Q = \Lambda_i(f_t^0 + f_{t-1}^0 + f_{t-2}^0)/3 + \epsilon_{i,q,t}^Q$.
- For monthly variables: $x_{i,m,t}^M = \Lambda_i f_t^0 + \epsilon_{i,m,t}^M$.

(iii) Obtain the initial values for $\Sigma_u = \text{diag}(\sigma_1^2, \cdots, \sigma_n^2)$ by using the resulting residual from the regressions above in the following way:
- For quarterly variables: $\sigma_i^2 = \frac{9}{3} V(\epsilon_{i,q,t}^Q)$.
- For monthly variables: $\sigma_i^2 = V(\epsilon_{i,m,t}^M)$.

(iv) Obtain the initial values for Φ by estimating a simple VAR(p) for the first r principal components f_t^0.

Alternative Variable Transformation and Aggregation Scheme for the DFM

Table A1 shows a comparison of the baseline model using three-month growth rates for the non-stationary variables and the alternative specification using month-on-month growth rates and the aggregation rule of Mariano and Murasawa (2003). In all cases, the baseline specification dominates the alternative version.

Table A1. Alternative Data Transformation and Aggregation Rules for DFM

	Early-Quarter Information Set		**Late-Quarter Information Set**	
	$h = 1$	$h = 2$	$h = 1$	$h = 2$
Baseline: Three-Month Growth Rates and 1/3, 1/3, 1/3 Aggregation	1.44	2.02	1.42	1.47
Alternative: Month-on-Month Growth Rates and 1/3, 2/3, 3/3, 2/3, 1/3 Aggregation	1.67	2.03	1.65	1.78

Notes: The table shows root mean squared errors (RMSEs) for $h = 1$ (nowcast) and $h = 2$ (one additional quarter ahead) using two different states of information. "Baseline" corresponds to the DFM model used in the paper and "Alternative" is the aggregation and transformation scheme proposed by Mariano and Murasawa (2003).

Model Combination Using MIDAS Regressions

As explained in the main text, we form the final forecast of the model combination approach by calculating the arithmetic average of all individual models that have a BIC that is 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). However, we experimented with various alternative combination schemes, which we briefly outline in the following.

Alternative Weighting Schemes

(i) Approximated Bayesian weighting scheme depending on the model's Bayesian (or Schwarz) information criterion. This scheme implies that $w_i^{SIC} = (\exp(-0.5 \cdot \Delta_i^{SIC}))/(\sum_{i=1}^{n} \exp(-0.5 \cdot \Delta_i^{SIC}))$ with $\Delta_i^{SIC} = SIC_i - SIC_{\min}$.

(ii) Mallows model averaging (MMA) as proposed by Hansen (2007, 2008). This measure is based on Mallows' criterion for model selection. The goal is to minimize the mean square error (MSE) over a set of feasible forecast combinations. This implies minimizing the function $C = (y - Fw)'(y - Fw) + w'Ks^2$, where K is a vector including the number of coefficients of each model and $s^2 = (\hat{e}(M)'\hat{e}(M))/(T - k(M))$ is an estimate for σ^2 from the model with the smallest estimated error variance. We apply the constraints $0 \leq w \leq 1$ and $\sum_{i=1}^{n} \omega_i = 1$. Note that MMA explicitly takes into account the number of estimated parameters of the model.

(iii) Weights that are inversely proportional to the models' past forecast errors as applied, for example, in Stock and Watson (2004) and Drechsel and Scheufele (2012) to predict output using leading indicators. For concreteness, consider the following example: t denotes the period with the last observation of the target. The weights applied to form the h-period-ahead forecast are calculated by recursively running out-of-sample forecasts for the last $p - 1$ periods and for

each indicator i. Based on the corresponding forecast errors $e_{i,t-p+1}, \cdots, e_{i,t}$, the mean square forecast error (MSFE) for each indicator i is calculated, $S^h_{i,t} = \sum_{s=t-p+1}^{t} \left(\hat{e}_{i,s}\right)^2 / p$. The weights that are then used to calculate the final forecast for $t+h$ GDP growth are

$$w^h_{i,t} = \frac{(S^h_{i,t})^{-1}}{\sum_{j=1}^{n} (S^h_{j,t})^{-1}}. \tag{8}$$

Note that the mean square forecast error and therefore also the weights are specific to the forecasting horizon h.

(iv) A rank-based weighting scheme. This scheme is also based on the models' past forecast performance; however, the weights are computed using the ranks of the individual models (Aiolfi and Timmermann 2006; Heinisch and Scheufele 2018). It is thus very closely related to the previous combination schemes that use the past MSFEs to calculate the weights. However, instead of relying directly on the MSFEs, the weights are calculated based on the models' ranks: we order the indicators according to the MSFEs, $S^h_{i,t}$, calculated under (iii). The model with the lowest $S^h_{i,t}$ obtains rank 1, $\mathcal{R}^h_{i,t} = f(S^h_{1,t}, \ldots, S^h_{n,t})$. The weights are then

$$w^h_{i,t} = \frac{(\mathcal{R}^h_{i,t})^{-1}}{\sum_{j=1}^{N} (\mathcal{R}^h_{j,t})^{-1}}. \tag{9}$$

One advantage of ranks compared with direct MSFE weights is that they are less sensitive to outliers and, thus, should be more robust. In practice, the weighting scheme based on ranks places very high weight on the group of best models and nearly zero weight on models with less-accurate past performance. Note that the mean square forecast error and therefore also the weights are specific to the forecasting horizon h.

Table A2. Alternative Forecast Combination Approaches

	Total Sample 2005:Q1–2015:Q2		**Post-Crisis Sample 2010:Q1–2015:Q2**	
	$h = 1$	$h = 2$	$h = 1$	$h = 2$
A. Early-Quarter Information				
Benchmark (Mean)	1.6660	2.1831	1.2411	1.1755
SIC	0.999	0.999	1.005*	0.999
MMA	1.140*	1.072	1.283**	1.134
MSE	0.995	1.007	1.037	1.010
RANK	0.982	0.990	1.158*	1.027
B. Late-Quarter Information				
Benchmark (Mean)	1.5604	2.0528	1.1498	1.1701
SIC	0.997	1.000	1.005	1.000
MMA	1.074	1.068	1.153	1.095
MSE	0.979	1.006*	1.028	1.011
RANK	0.969	0.965	1.115	0.995

Notes: The table shows 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 based on equal weights, all numbers are defined relative to the benchmark. SIC: weights based on information criterion; MMA: Mallows model averaging; MSE: inverse MSE of previous performance; RANK: inverse of the rank of the previous mean squared performance. ***, **, and * indicate whether a model's predictive ability (using the DM test) is significantly different from the benchmark (at the 1 percent, 5 percent, and 10 percent levels, respectively).

Empirical Performance of the Different Weighting Schemes. Table A2 presents the performance of the different weighting schemes relative to the baseline scheme. No alternative scheme yields a significant improvement relative to the arithmetic mean used in the baseline specification. However, the MMA weights clearly underperform the baseline specification.

Variable Selection Using a Specific-to-General Approach

In the following we describe various elements of the variable selection methodology in somewhat more detail.

Blocking and Realignment

In order to handle mixed frequency, we apply blocking. In particular, we generate for each monthly indicator $x^M_{i,m,t}$ three new quarterly indicators $\tilde{x}^Q_{i,q,t}$, $\tilde{x}^Q_{i,q,t-1}$, and $\tilde{x}^Q_{i,q,t-2}$ for $t = 3, 6, 9, \ldots$. This leaves us with a quarterly data set. However, this data set may still be unbalanced at the current margin due to differing publication lags. We therefore realign the indicator such that the panel becomes balanced.

The LM Statistic and the Critical Value with Which It Is Compared

The LM statistic for indicator k used in steps (iii) and (iv) of the algorithm described in the main text is calculated as $\lambda_k = TR^2_k$ (Godfrey 1988), where R^2_k is the coefficient of determination of the regression of step (iii).

The critical value with which we compare the highest LM statistic in step (v) is the $(1 - \tilde{\alpha})$ quantile of the $\chi^2(1)$ distribution. $\tilde{a}$ depends on the pre-specified significance level α, $\tilde{a} = f(\alpha)$. It is computed according to the false discovery rate control method (as proposed by Benjamini and Hochberg 1995). We use $\alpha = 0.05$.

Note that by setting $\tilde{a} \neq \alpha$, we deviate from the original method proposed by Herwartz (2011a, 2011b). In the case of large data sets, without any adjustment for multiple testing, the danger of overfitting is quite evident. However, by applying a standard Bonferroni adjustment, the probability of selecting an indicator would drop to zero when the size of the data set increases strongly. Therefore, we have to find a compromise between the two issues. The false discovery rate (FDR) achieves exactly that, as it was shown to have much better power properties than standard Bonferroni bounds and still guards against large type I errors.[2] Table A3 shows the performance of the Bonferroni correction and the original method applying no correction relative to our baseline. Bonferroni works similar to our correction, whereas applying no correction performs clearly worse for $h = 1$.

[2]More specifically, we employ the proposed two-step method of Benjamini, Krieger, and Yekutieli (2006).

Table A3. Modifications of the Specific-to-General Approach

	Total Sample 2005:Q1–2015:Q2		**Post-Crisis Sample 2010:Q1–2015:Q2**	
	$h = 1$	$h = 2$	$h = 1$	$h = 2$
A. Early-Quarter Information				
Benchmark (FDR)	1.5736	2.6475	1.5286	1.8642
Bonferroni	1.004	0.995	0.930	0.876
No Correction	1.182**	0.976	1.007	1.255
B. Late-Quarter Information				
Benchmark (FDR)	1.5739	2.6653	1.5303	1.8846
Bonferroni	1.004	0.992	0.931	0.866
No Correction	1.263**	0.978	1.064	1.253

Notes: The table shows 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 procedure, which uses the control of false discovery rate (FDR) of the significance level, all numbers are defined relative to the benchmark. Bonferroni: Bonferroni adjustment; No Correction: no correction for multiple testing. ***, **, and * indicate whether a model's predictive ability (using the DM test) is significantly different from the benchmark (at the 1 percent, 5 percent, and 10 percent levels, respectively).

References

Aiolfi, M., and A. Timmermann. 2006. "Persistence in Forecasting Performance and Conditional Combination Strategies." *Journal of Econometrics* 135 (1–2): 31–53.

Angelini, E., M. Bańbura, and G. Rünstler. 2010. "Estimating and Forecasting the Euro Area Monthly National Accounts from a Dynamic Factor Model." *OECD Journal: Journal of Business Cycle Measurement and Analysis* 2010 (1): 1–22.

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.

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.

Benjamini, Y., A. M. Krieger, and D. Yekutieli. 2006. "Adaptive Linear Step-up Procedures that Control the False Discovery Rate." *Biometrika* 93 (3): 491–507.

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.

Durbin, J., and S. J. Koopman. 2012. *Time Series Analysis by State Space Models.* 2nd Edition. Oxford University Press.

Godfrey, L. G. 1988. *Misspecification Tests in Econometrics.* Cambridge: Cambridge University Press.

Hansen, B. E. 2007. "Least Squares Model Averaging." *Econometrica* 75 (4): 1175–89.

———. 2008. "Least-Squares Forecast Averaging." *Journal of Econometrics* 146 (2): 342–50.

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.

Mariano, R., 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.

———. 2010. "A Coincident Index, Common Factors, and Monthly Real GDP." *Oxford Bulletin of Economics and Statistics* 72 (1): 27–46.

Rünstler, G., K. Barhoumi, S. Benk, R. Cristadoro, A. Den Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, and C. Van Nieuwenhuyze. 2009. "Short-term Forecasting of GDP Using

Large Datasets: A Pseudo Real-time Forecast Evaluation Exercise." *Journal of Forecasting* 28 (7): 595–611.

Stock, J. H., and M. W. Watson. 2004. "Combination Forecasts of Output Growth in a Seven-Country Data Set." *Journal of Forecasting* 23 (6): 405–30.