Weighted Maximum Likelihood for Dynamic Factor Analysis and Forecasting with Mixed Frequency Data

F. Blasques\textsuperscript{(a)}, S.J. Koopman\textsuperscript{(a,b)}, M. Mallee\textsuperscript{(a)} and Z. Zhang\textsuperscript{(a)*}

\textsuperscript{(a)} VU University Amsterdam & Tinbergen Institute Amsterdam
\textsuperscript{(b)} CREATES, Aarhus University, Denmark

September 30, 2015

Abstract

For the purpose of forecasting key macroeconomic or financial variables from a high-dimensional panel of mixed frequency time series variables, we adopt the dynamic factor model and propose a weighted likelihood-based method for the estimation of the typically high-dimensional parameter vector. The loglikelihood function is split into two parts that are weighted differently. The first part is associated with the key variables while the second part is associated with the related variables which possibly contribute to the forecasting of the key variables. We derive asymptotic properties, including consistency and asymptotic normality, of the weighted maximum likelihood estimator. We show that this estimator outperforms the standard likelihood-based estimator in approximating the true unknown distribution of the data as well as in out-of-sample forecasting accuracy. We also verify the new estimation method in a Monte Carlo study and investigate the effect of different weights in different scenarios. When forecasting gross domestic product growth, this key variable is typically observed at a low (quarterly) frequency while the supporting variables are observed at a high (monthly) frequency. For such a mixed frequency high-dimensional dataset, we adopt a low frequency dynamic factor model and discuss the computational efficiencies of this approach. In an extensive empirical study for the U.S. economy, we present improvements in forecast precision when the weighted likelihood-based estimation procedure is adopted.

Keywords: Asymptotic theory, Forecasting, Kalman filter, Nowcasting, State space.


1 Introduction

The forecasting of macroeconomic and financial time series variables is of key importance for economic policy makers. Reliable short-term forecasts are especially in high demand when the economic environment is uncertain as we have witnessed in the years during

\textsuperscript{*}S.J. Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. Emails: f.blasques@vu.nl, s.j.koopman@vu.nl, m.i.p.mallee@vu.nl and z2.zhang@vu.nl
and after the financial crisis that started in 2007. Many different model-based approaches exist for this purpose, ranging from basic time series models to sophisticated structural dynamic macroeconomic models. The underlying idea of the dynamic factor model is to associate a relatively small set of factors to a high-dimensional panel of economic variables that includes the variable of interest and related variables. The dynamic factor model has become a popular tool for the short-term forecasting of the variable of interest, amongst practitioners and econometricians. This is mainly due to their good forecast performance as shown in many studies.

The estimation of the parameters in a dynamic factor model is a challenging task given the high-dimensional parameter vector that mostly consists of factor loading coefficients. A likelihood-based approach in which the likelihood function is evaluated by the Kalman filter and numerically maximized with respect to the parameter vector is originally proposed by Engle and Watson (1981) for a model with one dynamic factor. Watson and Engle (1983) base their estimation procedure on the expectation maximization (EM) algorithm; see also Quah and Sargent (1993). Approximate likelihood-based procedures are developed by Doz, Giannone, and Reichlin (2011) and Baribura and Modugno (2014). Bräuning and Koopman (2014) and Jungbacker and Koopman (2015) propose efficient transformations of the model to facilitate parameter estimation for high-dimensional dynamic factor models. We restrict ourselves in this study to these likelihood-based estimation procedures.

Motivation

An important application of dynamic factor models is their use in the forecasting of growth in gross domestic product (GDP). A typically high-dimensional panel of macroeconomic variables is used to construct factors for the purpose of facilitating the forecasting of GDP growth. Empirical evidence is given by, amongst others, Stock and Watson (2002) and Giannone, Reichlin, and Small (2008) for the U.S., Marcellino, Stock, and Watson (2003) and Rünstler et al. (2009) for the euro area, and Schumacher and Breitung (2008) for Germany. In all these studies, the problem of mixed frequency data arises since the variable of interest GDP growth is observed at a quarterly frequency while the other macroeconomic variables are observed at a monthly frequency. The treatment of mixed frequency data in a dynamic factor model is therefore important for forecasting, nowcasting and backcasting GDP growth. The use of the dynamic factor model in this context is subject to two fundamental characteristics. First, the use of the factors for extracting commonalities in the dynamics in the variables is intended to provide a parsimonious way to link the variables of interest with the related variables. The dynamic factor model is not designed to provide a correct and exact representation of the true unknown data generating process for all variables in the panel. Instead, the factors provide a convenient tool to model relations that are potentially very complex. Second, the variable of interest and the related variables in the dynamic factor model are observed at different frequencies and play different roles. In particular, we wish to obtain accurate forecasts for
the variable of interest that is observed at a low frequency (quarterly) while the related high frequency (monthly or weekly) variables play a more secondary role as instruments to improve the forecasting accuracy of the key variable. These two fundamental characteristics of the dynamic factor model for macroeconomics play an important role in our parameter estimation procedure which is designed to establish a framework in which the standard maximum likelihood estimator can be improved upon. Next we discuss the two characteristics of the model in more detail.

**Weighted likelihood-based estimation**

To address the notion that a single or a small selection of variables in a dynamic factor model are of key interest while the others can be regarded as instruments, we present a weighted likelihood-based estimation procedure for the purpose of providing a more accurate forecasting performance than the standard maximum likelihood estimator. Our proposed weighted maximum likelihood estimator gives simply more weight to the likelihood contribution from the variable of interest. For example, for the nowcasting and forecasting of quarterly growth in gross domestic product, referred to as GDP growth, more weight can be given to likelihood contribution from GDP growth in comparison to the other, related variables in the dynamic factor model. The variable-specific weights introduced by this novel weighted ML estimator differ considerably from other weighted ML estimators proposed in the literature that introduce observation-specific weights in the likelihood function. The local ML estimators studied in Tibshirani and Hastie (1987), Staniswalis (1989) and Eguchi and Copas (1998) assign a weight to each observation that depends on the distance to a given fixed point. The robust ML estimator of Markatou, Basu, and Lindsay (1997, 1998) down-weights observations that are inconsistent with the postulated model. Similarly, Hu and Zidek (1997) devise a general principle of relevance that assigns different weights to different observations in an ML setting. In small samples, this type of estimator can provide important gains in the trade-off between bias and precision of the ML estimator. The large sample properties of these estimators are established in Wang, van Eeden, and Zidek (2004) for given weights, and Wang and Zidek (2005) provide a method for estimating the weights based on cross-validation. Contrary to these examples, we propose a weighted ML estimator that gives higher weight to a subset of a random vector, i.e. to an entire random scalar sequence within the modeled multivariate stochastic process. We discuss the asymptotic properties of this novel weighted maximum likelihood estimator and we show that the estimator is consistent and asymptotically normal. We further verify our new approach in a Monte Carlo study to investigate the effect of different choices for the weights in different scenarios. We also adopt the weighted likelihood function for parameter estimation in our empirical study concerning the nowcasting and forecasting of US GDP growth based on a full system dynamic factor model with mixed frequency variables.
Mixed Frequency

The common treatments of mixed frequency data are Bridge models and Mixed Data Sampling (MIDAS) models. In Bridge models the high frequency data are forecast up to the desired forecast horizon in a separate time series model. These forecasts are then aggregated to the lower frequency and are used as explanatory variables in a lower frequency time series model as contemporaneous values. Trehan (1989) is the first application of Bridge equations in our setting; see also Baffigi, Golinelli, and Parigi (2004) and Golinelli and Parigi (2007) for more recent uses of Bridge models. The MIDAS approach of Ghysels, Santa-Clara, and Valkanov (2006) regresses the low frequency time series variable of interest on the related high frequency variables; the dynamics of the regressors are not considered by the model. The high frequency regressors are not aggregated but each individual lag has its own regression coefficient. To avoid parameter proliferation, these coefficients are subject to specific weighting functions. Foroni, Marcellino, and Schumacher (2015) propose the use of unconstrained distributed lags and refer to this approach as unrestricted MIDAS (U-MIDAS).

Dynamic factor models and vector autoregressive (VAR) models can also be adapted to handle mixed frequency data. The study of Mariano and Murasawa (2003) is the first to illustrate how a small-scale dynamic factor model can be adapted for mixed frequency data. The model is formulated in state space form with a monthly time index. The monthly and quarterly variables are then driven by a single monthly dynamic factor and by an idiosyncratic factor for each individual series. The Kalman filter treats the missing observations in the quarterly which occur during the first two months of the quarter. VAR models for mixed frequency time series has originally proposed by Zadrozny (1990). In this approach the model also operates at the highest frequency in the data and variables which are observed at a lower frequency are viewed as being periodically missing. Mittnik and Zadrozny (2005) report promising results based on this original approach for the forecasting of German growth in gross domestic product. Ghysels (2012) generalizes the MIDAS approach for the mixed frequency VAR (MF-VAR). Schorfheide and Song (2013) develop a MF-VAR model that allows some series to be observed at monthly and others at quarterly frequencies, using Bayesian state space methods. Their empirical findings for a large-scale economic VAR are highly promising: within-quarter monthly information leads to drastic improvements in short-horizon forecasting performance. Bai, Ghysels, and Wright (2011) examine the relationship between MIDAS regressions and state space models applied to mixed frequency data. They conclude that the Kalman filter forecasts are typically slightly better, but MIDAS regressions can be more accurate if the state space model is misspecified or over-parameterized. Kuzin, Marcellino, and Schumacher (2011) compare the accuracy of Euro Area GDP growth forecasts from MIDAS regressions and MF-VARs estimated by maximum likelihood.

Marcellino, Carriero, and Clark (2014) propose a Bayesian regression model with stochastic volatility for producing current-quarter forecasts of GDP growth with a large range of available within-the-quarter monthly observations of economic indicators. Each
time series of monthly indicators is transformed into three quarterly time series, each containing observations for the first, second and third month of the quarter. We adopt their overall approach but apply it to the dynamic factor model and formalize it within the state space modeling framework in a mixed frequency data setting. Similar ideas are developed for periodic systems in the control engineering literature; see Bittanti and Colaneri (2000, 2009). In the econometrics literature and for vector autoregressive systems, such ideas are explored in Chen, Anderson, Deistler, and Filler (2012), Ghysels (2012), Foroni, Gurin, and Marcellino (2015) and Ghysels, Hill, and Motegi (2015).

Outline

The outline of the paper is as follows. In Section 2 we present our weighted maximum likelihood approach that is introduced to increase the influence of the key variables in the estimation process for a joint multivariate dynamic model. Asymptotic properties of the resulting estimator are derived and we explore its small-sample properties in a Monte Carlo study. In Section 3 we show how mixed frequency dynamic factor models can be specified as observationally equivalent low frequency dynamic factor models. In many cases the new formulations lead to computational gains. In Section 4 we present and explore the results of our empirical study concerning US GDP growth. We compare the nowcasting and forecasting accuracies of our new approach with those of a benchmark model. We also establish the empirical relevance of the weighted estimation method of Section 2. Section 5 summarizes and concludes.

2 Weighted Maximum Likelihood Estimation: method and properties

We represent our high-dimensional unbalanced panel of time series as the column vector $z_t$ for which we have observations from $t = 1, \ldots, T$ where $T$ is the overall time series length. An unbalanced panel refers to the case where we have observations for some of the variables in the vector $z_t$. We assume that at least some observations are available for each variable in $z_t$, for $t = 1, \ldots, T$. The entries for which no observations are available in $z_t$ are treated as missing at random. We decompose $z_t$ into variables of interest in $y_t$ and related variables in $x_t$, we have $z_t = (y_t', x_t')'$ where $a_t'$ is the transpose of column vector $a_t$. The dimension $N_y$ of $y_t$ is small and typically equal to one while the dimension $N_x$ of $x_t$ can be large. Hence the dimension $N_z$ of $z_t$ is also large since $N_z = N_y + N_x$. It is assumed that all time series variables in $z_t$ have zero mean and are strictly stationary. The basic dynamic factor model for $z_t$ can be represented by $z_t = \Lambda f_t + \varepsilon_t$ or

$$
\begin{pmatrix}
  y_t \\
  x_t
\end{pmatrix}
 = \begin{bmatrix}
  \Lambda_y \\
  \Lambda_x
\end{bmatrix} f_t + \begin{pmatrix}
  \varepsilon_{y,t} \\
  \varepsilon_{x,t}
\end{pmatrix}, \quad f_t = \Phi f_{t-1} + \eta_t,
$$

(1)
for $t = 1, \ldots, T$, where $\Lambda = [\Lambda_y', \Lambda_x']'$ is the factor loading matrix with dimension $N_z \times r$, with $A'$ being the transpose of matrix $A$, $f_t$ is the $r \times 1$ vector with latent dynamic factors, $\varepsilon_t = (\varepsilon_{yt}, \varepsilon_{xt})'$ is the normally distributed observation disturbance vector, $\Phi$ is the autoregressive coefficient matrix, and $\eta_t$ is the normally distributed factor disturbance vector. The dynamic factor $f_t$ represents the common dynamic variations in the time series variables in $z_t$. The dynamic process for $f_t$ is specified as a strictly stationary vector autoregressive process. Hence, matrix $\Phi$ is subject to the appropriate conditions for stationarity. Other stationary, linear dynamic processes can also be considered for $f_t$. For identification purposes we further assume that the factors are normalized, that is $\mathbb{E}(f_t) = 0$ and $\text{Var}(f_t) = I_r$ with $I_k$ being the $k \times k$ identity matrix for any positive integer $k$. In our treatment below, the initial factor $f_1$ is treated as a fixed value, that is $f_1 = f_1^*$. The disturbance vectors $\varepsilon_t$ and $\eta_t$ are assumed to be mutually and serially uncorrelated for all time periods. In particular, we have

$$\varepsilon_t \sim N(0, \Sigma_\varepsilon), \quad \eta_t \sim N(0, \Sigma_\eta), \quad \text{Cov}(\varepsilon_t, \eta_s) = 0,$$

for $t, s = 1, \ldots, T$. To enforce the normalization of the factors $f_t$, the variance matrix $\Sigma_\eta$ is restricted to be $\Sigma_\eta = I_r - \Phi \Phi'$. The remaining coefficient matrices $\Lambda$, $\Sigma_\varepsilon$ and $\Phi$ are functions of the unknown parameter vector that we denote by $\psi$. This dynamic factor model is stylized for the purpose of presenting our developments below. However, our results are general for other multivariate dynamic specifications, including those for mixed frequencies. We will discuss these generalizations in the following sections.

Different methods have been proposed for the estimation of the unknown parameter vector $\psi$; see the discussion in the introductory section. We restrict ourselves to those methods that are likelihood-based and aim at the maximization of the loglikelihood function. The loglikelihood function relies on the joint logdensity $\log p(z; \psi) = \log p(y, x; \psi)$ where $p(\cdot)$ is the Gaussian density function, $z = (y', x')'$, $y = (y_1', \ldots, y_T')'$ and $x = (x_1', \ldots, x_T')'$. Since the dynamic factor model (1)-(2) can be represented as a stationary Gaussian linear state space model, the Kalman filter can be used to evaluate the loglikelihood function; we refer to Harvey (1989) and Durbin and Koopman (2012) for treatments of the Kalman filter. The maximum likelihood (ML) estimates are obtained by numerically maximizing the loglikelihood function with respect to $\psi$. It is a standard exercise of numerical optimization in which the Kalman filter evaluates the loglikelihood function whenever a different $\psi$ is considered. However, the ML estimator is not necessarily the best estimator in the context of the dynamic factor model for the following two reasons: (i) the dynamic factor model only provides a parsimonious approximation to a high-dimensional complex data generation process of the variables in $z_t$; (ii) the dynamic factor model is typically used for the forecasting of the variables in $y_t$ rather than the forecasting of all variables in $z_t$. The dynamic factor model provides a convenient framework for obtaining simple descriptions of potentially complex interactions between the economic variables. In particular, the common factors summarize partly the com-
monalities in the dynamic variations in the related variables $x_t$. Furthermore, the factors deliver a parsimonious description of the relationships between the variables of interest in $y_t$ and the related variables in $x_t$. The dynamic factor model is mainly used to approximate the true and unknown data generation process. It is not intended to be an exact representation of the true underlying dynamics of the economy.

In the particular context of forecasting GDP growth as discussed in the introductory section, we are concerned with the asymmetric treatment of the variables in the mixed frequency dynamic factor model. Specifically, while the low frequency variable plays the role of ‘variable of interest’, the high frequency variables play only the role of instruments that are aimed to improve the forecasting accuracy of the low frequency variable. In other words, all efforts lie on improving the forecasting accuracy of the low frequency variable, as opposed to approximating the joint distribution of the observed data as a whole.

Hence, in the context of parameter estimation, we need to address the problem of model misspecification and the focus on a subset of variables only. Each of these issues are not necessarily sufficient to abandon the ML estimator, but taken together, they are. Indeed, if we are only interested in a subset of the variables, but the model is correctly specified, then the ML estimator is still the best under the usual regularity conditions that make it consistent and efficient. In particular, by converging to the true parameter and attaining a minimum variance, the ML estimator provides the best possible parameter estimates for the purpose of forecasting the variable of interest. This is true even if the variable of interest happens to be only a subset of the observed variables. Similarly, if a model is misspecified but our interest lies in forecasting all observed variables, then there are still very favorable reasons to employ the ML estimator. Under well-known conditions, the ML estimator converges to a pseudo-true parameter that minimizes the Kullback-Leibler (KL) divergence between the true joint distribution of the data and the model-implied distribution. The KL divergence has well established information-theoretic optimal properties. Furthermore, under weak regularity conditions and depending on the distribution of the data, it is also easy to show that the limit pseudo-true parameter optimizes forecasting accuracy. As we shall see below however, when taken together, the above points (i) and (ii) imply that the ML estimator is no longer the best possible estimator available. As such, these two features that characterize our dynamic factor model call for a novel estimation approach that improves the forecasting accuracy of the variable of interest. We provide both theoretical and simulation-based evidence that a weighted ML estimator outperforms the classic ML estimator in forecasting the variable of interest. The ability to outperform the ML estimator is also visible in empirically relevant applications to economic data.

In this section we consider the basic dynamic factor model (1) and (2) where both $y_t$ and $x_t$ can be treated as vectors. These development can be adapted easily for the mixed frequency dynamic factor model of the next section and other more general specifications.
The loglikelihood function for the model can be given by

$$L_T(\psi, f_1^*) := \log p(y, x; \psi) = \log p(y|x; \psi) + \log p(x; \psi),$$

(3)

where the initial value of the factor $f_1 = f_1^*$ and the parameter vector $\psi$ are both treated as fixed unknown values. It is a standard result that the joint density can be expressed as a conditional density multiplied by a marginal density. However, for our purposes the expression (3) is useful as it highlights the different roles of $y$ and $x$: the variable $y_t$ is our key variable for which we require accurate model-based forecasts while the variables represented by $x_t$ are typically instrumental to improve the nowcasts and forecasts of $y_t$. Under the assumption that $y$ and $x$ are jointly generated by the Gaussian dynamic factor model (1), we can apply the Kalman filter to evaluate the loglikelihood function via the prediction error decomposition. Koopman and Durbin (2000) discuss an alternative filtering method in which each element of the observation vector $(y', x')'$ is brought one at a time into the updating step of the Kalman filter. The vector series $z_t$ is effectively converted into a univariate time series where multiple observations are available for the same time index.

The maximum likelihood estimation of parameter vector $\psi$ is based on applying a numerical quasi-Newton optimization method for the maximization of $L_T(\psi, f_1^*)$, with respect to $\psi$. The maximization is an iterative process. After convergence, the maximum likelihood estimate of $\psi$ is obtained. For each iteration in this process, various loglikelihood evaluations are required and they are carried out by the Kalman filter. In the context of the mixed frequency dynamic factor model, the treatment of the observations in $z_t$ for the construction of the likelihood function is implied by the dynamic factor model. However, it is very likely that the dynamic factor model is misspecified as a model representation of the true data generation process for the variables represented in $z_t$. When our primary aim is to analyze $y_t$ in particular, we may be less concerned with the misspecification of $x_t$, to some extent. To reflect the higher importance of $y_t$ in comparison to $x_t$ in the likelihood construction for the misspecified dynamic factor model, we propose to give different weights to the likelihood contributions of $y_t$ and $x_t$ explicitly. Hence we propose the weighted loglikelihood function

$$L_T(\psi, w, f_1^*) = W \cdot \log p(y|x; \psi) + \log p(x; \psi),$$

(4)

for a fixed and predetermined weight $W \geq 1$ and with $w := W^{-1} \in [0, 1]$. The weight $W$ is conveniently used in our Monte Carlo and empirical studies below while it is more appropriate to work with the inverse weight $w$ in the asymptotic theory that is developed next. The construction of the weighted loglikelihood function does not need further modifications. The estimator of $\psi$ that maximizes (4) is referred to as the weighted maximum likelihood (WML) estimator.

This novel WML estimator differ considerably from other weighted ML estimators proposed in the literature. To our knowledge, this WML estimator is unique in introducing
variable-specific weights rather than observation-specific weights in the likelihood function. For example, local ML estimators assign a weight to each observation that depends on the distance to a given fixed point; see Tibshirani and Hastie (1987), Staniswalis (1989) and Eguchi and Copas (1998). The robust ML estimator of Markatou, Basu, and Lindsay (1997, 1998) are designed to reduce influence of outliers by down-weighting observations that are inconsistent with the postulated model. The general principle of relevance of Hu and Zidek (1997) assigns different weights to different observations in the likelihood function.

The motivation for the development of our WML estimator is also considerably different. Our WML estimator is designed to perform well when the model is misspecified and interest lies in forecasting only a subset of the observed variables. For this reason we analyze the asymptotic properties of our WML estimator allowing for the possibility of model misspecification and focus on the approximation to an unknown data generation process. In contrast, the motivation for the weighted ML estimators found in the literature is typically related to gains in the trade-off between bias and precision of the ML estimator in the standard case of correct specification. As Wang, van Eeden, and Zidek (2004) derive asymptotic properties for observation-specific weighted ML estimators in the standard context if correct specification.

2.1 Asymptotic Properties of the WML Estimator

The properties of the weighted maximum likelihood estimator are derived for any choice of weight \( w := W^{-1} \in [0, 1] \). We show that, when the model is correctly specified, then the WML estimator \( \hat{\psi}_T(w) \) is consistent and asymptotically normal for the true parameter vector \( \psi_0 \in \Psi \). When the model is misspecified, we show that \( \hat{\psi}_T(w) \) is consistent and asymptotically normal for a pseudo-true parameter \( \psi^*_0(w) \in \Psi \) that minimizes a transformed Kullback–Leibler (KL) divergence between the true probability measure of the data and the measure implied by the model. We show that the transformed KL divergence takes the form of a pseudo-metric that gives more weight to fitting the conditional density of \( y_t \) when \( 0 < w < 1 \). For the special case where \( w = 1 \), we obtain the classical pseudo-true parameter \( \psi^*_0(1) \in \Psi \) of the ML estimator that minimizes the KL divergence. The proofs of all theorems presented in this section can be found in Appendix B.

Proposition 1 below states well known conditions for the strict stationarity and ergodicity (SE) of the true processes \( \{f_t\}_{t \in \mathbb{Z}}, \{x_t\}_{t \in \mathbb{Z}} \) and \( \{y_t\}_{t \in \mathbb{Z}} \) generated by the linear Gaussian model in (1) and (2), initialized in the infinite past.

**PROPOSITION 1.** Let \( \{x_t\}_{t \in \mathbb{Z}} \) and \( \{y_t\}_{t \in \mathbb{Z}} \) be generated according to (1) and (2) with

1. \( \|\Phi\| < 1 \) in (1) and \( 0 < \|\Sigma_\eta\| < \infty \) in (2);
2. \( \|\Lambda_x\| < \infty \) in (1) and \( 0 < \|\Sigma_\epsilon\| < \infty \) in (2);
3. \( \|\Lambda_y\| < \infty \) in (1).
Then \( \{x_t\}_{t \in \mathbb{Z}} \) and \( \{y_t\}_{t \in \mathbb{Z}} \) are SE sequences with bounded moments of any order; i.e. 
\( \mathbb{E}|x_t|^r < \infty \) and \( \mathbb{E}|y_t|^r < \infty \) \( \forall r > 0 \).

Theorem 1 ensures the existence of the WML estimator as a random variable that takes values in the arg max set of the random likelihood function.

**Theorem 1. (Existence)** For given \( w \in [0, 1] \), let \((\Psi, \mathfrak{B}(\Psi))\) be a compact measurable space. Then there exists a.s. a measurable map \( \tilde{\psi}_T(w, \tilde{f}_1^*) : \Omega \to \Psi \) satisfying

\[
\tilde{\psi}_T(w, \tilde{f}_1^*) \in \arg \max_{\psi \in \Psi} L_T(\psi, w, \tilde{f}_1^*),
\]

for all \( T \in \mathbb{N} \) and every filter initialization \( \tilde{f}_1^* \).

Theorem 2 establishes the strong consistency of the WML estimator of the true parameter vector \( \psi_0 \in \Psi \) for any choice of weight \( w \in (0, 1] \) for the likelihood. This result is obtained under the assumption that the mixed frequencies common factor model is well-specified and for any filter that identifies the parameter vector \( \psi_0 \in \Psi \) and is asymptotically SE with bounded moments of second order. The identification conditions and exponential almost sure (e.a.s.) convergence of different filters to an SE process with bounded second moment is well known and easy to establish in this linear Gaussian setting. For this reason, we do not repeat them here; see e.g. Mehra (1970) for such results on the classical Kalman filter, Bougerol (1993) for extensions, and Blasques, Koopman, and Lucas (2014) for identification, convergence results and bounded moments on a wide range of observation-driven filters. Theorem 2 thus assumes that \( \psi_0 \) maximizes the likelihood and assumes the convergence of the filtered sequence \( \{\tilde{f}_t(\tilde{f}_1^*)\}_{t \in \mathbb{N}} \) initialized at \( \tilde{f}_1^* \) to a unique limit SE sequence \( \{\tilde{f}_t\}_{t \in \mathbb{Z}} \) with bounded second moment. Notice that we just require identification in the usual ML setting \( w = 1 \); i.e. identification w.r.t. the unweighted likelihood function \( L_T(\psi, 1) \). As shown in the proof, identification of \( \psi_0 \) in \( L_T(\psi, 1) \) implies identification of \( \psi_0 \) in \( L_T(\psi, w) \) for any \( w \in (0, 1] \).

**Theorem 2. (Consistency)** Let \( \{x_t\} \) and \( \{y_t\} \) be generated by the dynamic factor model defined in (1) and (2) under some \( \psi_0 \in \Psi \), and suppose that the conditions of Propositions 1 and Theorem 1 hold. Suppose furthermore that

\[
L_\infty(\psi_0, 1) > L_\infty(\psi, 1) \quad \forall \psi \neq \psi_0
\]

and there exists a unique SE sequence such that

\[
||\tilde{f}_t(\tilde{f}_1^*) - \tilde{f}_t|| \to^a 0 \quad \forall \tilde{f}_1^* \quad \text{as } t \to \infty \quad \text{with} \quad \mathbb{E}||\tilde{f}_t||^2 < \infty.
\]

Then the WML estimator \( \hat{\psi}_T(w, \tilde{f}_1^*) \) satisfies

\[
\hat{\psi}_T(w, \tilde{f}_1^*) \xrightarrow{a.s.} \psi_0 \quad \text{as } T \to \infty
\]

for any choice of weight \( w \in (0, 1] \) and any initialization \( \tilde{f}_1^* \).
If the data \( \{x_t\} \) and \( \{y_t\} \) are obtained from an unknown data generating process but satisfy some regularity conditions, then we can still prove consistency of the WML estimator to pseudo-true parameter \( \psi_0(w) \in \Psi \) that depends on the choice of weight \( w \in (0, 1] \).

It is well known that the classical ML estimator converges to a limit pseudo-true parameter that minimizes the KL divergence between the true joint probability measure of the data and the measure implied by the model. Theorem 3 characterizes the limit pseudo-true parameter \( \psi_0(w) \) as the minimizer of a transformed KL divergence for every given \( w \in (0, 1] \). Similar to the KL divergence, this new transformed divergence is also a pre-metric on the space of probability measures. The transformed KL divergence is further shown to be a weighted average of two KL divergences that is bounded from above (for \( w = 1 \)) by the KL divergence of the joint density of \( y_t \) and \( x_t \), and bounded from below (for \( w = 0 \)) by the conditional density of \( y_t \) given \( x_t \). For \( w \in (0, 1) \) the WML estimator converges to a pseudo-true parameter that gives more weight to the fit of the conditional model for \( y_t \) than the standard ML estimator.

Below we let \( p \) denote the true joint density of the vector \( z_t := (y_t, x_t)' \), where \( x_t \) is the stacked vector of monthly variables \( x_t \), and let \( p(z_t) = p_1(y_t | x_t) \cdot p_2(x_t) \) so that \( p_1 \) denotes the true conditional density and \( y_t \) given \( x_t \) and \( p_2 \) the true marginal of \( x_t \). Similarly, we let \( q(\cdot; \psi) \) denote the joint density of \( z_t \) as defined by our parametric model under \( \psi \in \Psi \), and let \( q_1(\cdot; \psi) \) and \( q_2(\cdot; \psi) \) be the counterparts of \( p_1 \) and \( p_2 \) for the parametric model density. Finally, given any two densities \( a \) and \( b \), we let \( \text{KL}(a, b) \) denote the KL divergence between \( a \) and \( b \).

**Theorem 3.** (Consistency) Let \( \{x_t\} \) and \( \{y_t\} \) be SE and satisfy \( \mathbb{E}|x_t|^2 < \infty \) and \( \mathbb{E}|y_t|^2 < \infty \). Furthermore, let the conditions of Theorem 1 hold and suppose that

\[
\mathcal{L}_\infty(\psi_0(w), w) > \mathcal{L}_\infty(\psi, w) \quad \forall \ \psi \neq \psi_0(w)
\]

and there exists a unique SE sequence such that

\[
\|\hat{f}_t(\hat{f}^*_1) - \hat{f}_t\| \overset{e.a.s.}{\to} 0 \quad \forall \ \hat{f}^*_1 \quad \text{as} \quad t \to \infty \quad \text{with} \quad \mathbb{E}||\hat{f}_t||^2 < \infty.
\]

Then

\[
\psi_T(w, \hat{f}^*_1) \overset{a.s.}{\to} \psi_0(w) \quad \text{as} \quad T \to \infty
\]

for any initialization \( \hat{f}^*_1 \) and any weight \( w \in (0, 1] \). Furthermore, the pseudo-true parameter \( \psi_0(w) \) minimizes a transformed KL divergence

\[
\text{TKL}_w(q(\cdot; \psi), p) = \text{KL}(q_1(\cdot; \psi), p_1) + w\text{KL}(q_2(\cdot; \psi), p_2)
\]

which is a pre-metric on the space of distributions satisfying for any \( w \in (0, 1] \),

\[
\text{TKL}_1(q(\cdot; \psi), p) = \text{KL}(q(\cdot; \psi), p), \quad \text{TKL}_0(q(\cdot; \psi), p) = \text{KL}(q_1(\cdot; \psi), p_1),
\]

11
Theorem 4 establishes the asymptotic normality of the WML estimator under the assumption that the mixed frequencies dynamic factor model is well specified. Below we let $J(\psi_0, w) := \mathbb{E}^{\psi_0}_{\psi_0, w}(\psi_0, w)^T$ denote the expected outer product of gradients and $I(\psi_0, w) := \mathbb{E}^{\psi_0}_{\psi_0, w}(\psi_0, w)$ be the Fisher information matrix. The asymptotic normality proof is written for filters whose derivative processes are asymptotically SE and have bounded moments; see Blasques et al. (2014) for a wide range of observation-driven filters satisfying such conditions. Below, $\{\tilde{df}_i(\tilde{df}_i^*)\}$ and $\{d\tilde{df}_i(\tilde{df}_i^*)\}$ denote the first and second derivatives of the filter w.r.t. the parameter vector $\psi$, initialized at $\tilde{df}_i^*$ and $d\tilde{df}_i^*$, respectively. Their SE limits are denoted $\{\tilde{df}_i\}$ and $\{d\tilde{df}_i\}$. Note that asymptotic normality result holds for any weight $w \in (0, 1]$, but the asymptotic distribution of the WML estimator depends on the choice of weight $w$.

**Theorem 4. (Asymptotic Normality)** Let the conditions of Theorem 2 hold and $\psi_0$ be a point in the interior of $\Psi$. Suppose furthermore that there exists a unique SE sequence $\{\tilde{df}_i\}$ such that

$$\|\tilde{df}_i(\tilde{df}_i^*) - \tilde{df}_i\| \xrightarrow{\mathcal{C}} 0 \quad \forall \tilde{df}_i \quad \text{as} \quad t \to \infty \quad \text{with} \quad \mathbb{E}|\tilde{df}_i|^4 < \infty$$

and a unique SE sequence $\{d\tilde{df}_i\}$ such that

$$\|d\tilde{df}_i(d\tilde{df}_i^*) - d\tilde{df}_i\| \xrightarrow{\mathcal{C}} 0 \quad \forall d\tilde{df}_i \quad \text{as} \quad t \to \infty \quad \text{with} \quad \mathbb{E}|d\tilde{df}_i|^2 < \infty.$$

Then, for every $\tilde{f}_i^*$ and every $w \in (0, 1]$, the ML estimator $\hat{\psi}_T(\tilde{f}_i^*)$ satisfies

$$\sqrt{T}(\hat{\psi}_T(\tilde{f}_i^*, w) - \psi_0) \xrightarrow{d} N\left(0, \ I^{-1}(\psi_0, w)J(\psi_0, w)J(\psi_0, w)\right) \quad \text{as} \quad T \to \infty.$$

Naturally, we can extend the asymptotic normality results to the mis-specified mixed measurement dynamic factor model by centering the WML estimator at the pseudo-true parameter $\psi_0^*(w)$.

**Theorem 5. (Asymptotic Normality)** Let the conditions of Theorem 3 hold and $\psi_0^*(w)$ be a point in the interior of $\Psi$. Suppose further that $\{x_t\}$ and $\{y_t\}$ are SE and satisfy $\mathbb{E}|x_t|^4 < \infty$ and $\mathbb{E}|y_t|^4 < \infty$ and there exists a unique SE sequence $\{\tilde{df}_i\}$ such that

$$\|\tilde{df}_i(\tilde{df}_i^*) - \tilde{df}_i\| \xrightarrow{\mathcal{C}} 0 \quad \forall \tilde{df}_i \quad \text{as} \quad t \to \infty \quad \text{with} \quad \mathbb{E}|\tilde{df}_i|^4 < \infty$$

and a unique SE sequence $\{d\tilde{df}_i\}$ such that

$$\|d\tilde{df}_i(d\tilde{df}_i^*) - d\tilde{df}_i\| \xrightarrow{\mathcal{C}} 0 \quad \forall d\tilde{df}_i \quad \text{as} \quad t \to \infty \quad \text{with} \quad \mathbb{E}|d\tilde{df}_i|^2 < \infty.$$
Then, for every \( \tilde{f}_1^* \) and every \( w \in (0, 1] \), the ML estimator \( \hat{\psi}_T(w, \tilde{f}_1^*) \) satisfies
\[
\sqrt{T}(\hat{\psi}_T(\tilde{f}_1^*) - \psi_0^*(w)) \xrightarrow{d} N\left(0, \mathcal{I}^{-1}(\psi_0^*(w), w)\mathcal{J}(\psi_0^*(w), w)\mathcal{I}^{-1}(\psi_0^*(w), w)\right) \quad \text{as} \quad T \to \infty.
\]

2.2 Selecting Optimal Weights

In this section we follow Wang and Zidek (2005) in proposing a method for estimating optimal weights that is based on cross-validation. In particular, we will focus on obtaining weights that optimize the out-of-sample forecasting performance of the low frequency variable of interest. Furthermore, we propose the use of a Diebold-Mariano test that allows us to infer if the improvements in forecasting accuracy produced by different choices of weights are statistically significant; see Diebold and Mariano (1995). We confirm the validity of the asymptotic distribution of the Diebold-Mariano test statistic under our set of assumptions.

For the purpose of estimating \( w \) by cross-validation, we will split the sample in two parts. The first part of the sample is used to estimate the model parameters, for any given choice of \( w \). The second part of the sample is used to evaluate the out-of-sample forecast performance of the model and select the optimal weight \( w \). Specifically, for some given \( w \), we first estimate the parameter vector \( \hat{\psi} \) using observations from period \( t = 1 \) to \( t = T' \). The parameter estimate, denoted \( \hat{\psi}_{1:T'}(w, \tilde{f}_1^*) \), is used to produce a one-step-ahead prediction \( \hat{y}_{T'+1}(\hat{\psi}_{1:T'}(w, \tilde{f}_2^*)) \) for the related variable. Next, we obtain an estimate \( \hat{\psi}_{2:T'+1}(w, \tilde{f}_2^*) \) using observations from period \( t = 2 \) to \( t = T' + 1 \) and produce another one-step-ahead prediction \( \hat{y}_{T'+2}(\hat{\psi}_{2:T'+1}(w, \tilde{f}_2^*)) \). We repeat this procedure and obtain \( H = T - T' - 1 \) one-step-ahead predictions using recursive samples, each based on the previous \( T' \) observations, as illustrated below,

\[
\begin{array}{cccccccc}
  y_1 & y_2 & y_3 & \cdots & \cdots & y_{T'} & \hat{y}_{T'+1} & \hat{y}_{T'+2} \\
  y_2 & y_3 & \cdots & \cdots & y_{T'} & y_{T'+1} & \cdots & \cdots \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
  y_H & \cdots & y_{T'} & y_{T'+1} & y_{T'+2} & \cdots & y_{T'+H} & \hat{y}_{T'+H+1}. \\
\end{array}
\]

If the WML estimator is well defined, then the one-step-ahead forecasts can effectively be written as a function of \( w \) since the WML estimator \( \hat{\psi}_{2:T'+1} \) maps every weight \( w \) to a point in the parameter space that defines a forecast value \( \hat{y}_{T'+i}(w) \equiv \hat{y}_{T'+i}(\hat{\psi}_{i:T'+i}(w)) \).

Finally, we define the \( H \) out-of-sample one-step-ahead forecast errors as follows
\[
e_i(w) = \hat{y}_{T'+i}(w) - y_{T'+i}, \quad i = 1, \ldots, H
\]
and use these to obtain a cross-validation criteria for selecting the weight \( w \) that minimizes the one-step-ahead mean squared forecast error (MSFE\(_1\)(w))
\[
\hat{w}_H = \arg\min_{w \in [0, 1]} \frac{1}{H} \sum_{i=1}^{H} e_{i+1}^2(w) = \arg\min_{w \in [0, 1]} \text{MSFE}_1(w)
\]
Naturally, the criterion can be easily redesigned for \( w \) to minimize the \( h \)-step-ahead forecast error (MSFE\(_h\)). Since \( w \) is directly chosen to minimize the forecast error, it is clear that any estimate \( \hat{w}_H \neq 1 \) will only occur if the WML estimator can improve the error compared to the ML estimator. However, it is important to take into account the possibility of spurious reductions in the MSFE that occur only because \( H \) is small. For this reason we propose the use of a Diebold-Mariano test statistic that can be used to assess if the improvement in forecasting accuracy is statistically significant. Lemma 1 highlights that the asymptotic Gaussian distribution derived in Diebold and Mariano (1995) is valid under the conditions of Theorem 5. The assumptions of the Diebold-Mariano test hold in the current setting, for any given pair \( (w, w') \), since Theorem 5 ensures that the data is SE with four bounded moments. The squared residuals are therefore covariance stationary and so are their differences. Of course, the question whether these assumptions hold in practice is ultimately an empirical issue for which there exist tests that one may wish to employ; see also the discussion in Diebold (2012).

Next we let \( \bar{d}_H(w, w') \) and \( \Sigma_H(d_i(w, w')) \) denote the sample average and standard error based on \( H \) differences in MSFE obtained under the weights \( w \) and \( w' \),

\[
d_i(w, w') := e_i(w)^2 - e_i(w')^2 \quad i = 1, \ldots, H.
\]

**Lemma 1.** Let the conditions of Theorem 5 hold. Then

\[
\bar{d}_H(w, w')/\Sigma_H(d_i(w, w')) \xrightarrow{d} N(0, 1) \quad \text{as} \quad H \to \infty
\]

under the null hypothesis \( H_0 : \mathbb{E}d_i(w, w') = 0 \), and

\[
\bar{d}_H(w, w')/\Sigma_H(d_i(w, w')) \to \infty \quad \text{as} \quad H \to \infty
\]

under the alternative hypothesis \( H_1 : \mathbb{E}d_i(w, w') > 0 \).

We stress that the Diebold-Mariano test is the most natural tool for comparing the forecasting performance of our model under any two WML estimates. The more recent tests proposed in the literature, for example, West (1996) and Clark and McCracken (2001, 2015) are not appropriate for our comparisons as they focus on testing the forecasting performance of different models evaluated at their pseudo-true parameters, rather than testing different forecasts; see also the discussions in Giacomini and White (2006) and Diebold (2012).

### 2.3 Small Sample Properties of WML: A Monte Carlo Study

Next we investigate the finite sample effects of different choices for the value of \( W \) on the in-sample fit in different settings of the data generation process (DGP) using Monte Carlo simulations. We generate data for a univariate time series \( y_t, N_y = 1 \), and a \( N_x \times 1 \) vector time series \( x_t \), for different panel dimensions \( N_x = 2, N_x = 5 \) and \( N_x = 10 \). The
length of the time series is set to $T = 120$ for all cases. The window length of the sample for parameter estimation is 80. The remaining 40 observations are used for out-of-sample forecast evaluation and the selection of the optimal value for $W$. We consider two different data generating processes for the vector of observations $z_t = (y_t, x'_t)'$ in the simulations.

The first DGP is (1) with one common factor $f_t$, $r = 1$, but with the addition of an idiosyncratic autoregressive component to each variable. The model is then given by

$$ z_t = \Lambda f_t + u_t + \varepsilon_t, \quad u_t = Bu_{t-1} + \xi_t, \quad \xi_t \sim N(0, \Sigma_\xi), $$

(5)

for $t = 1, \ldots, T$, where $B$ is a diagonal matrix with diagonal elements $-1 < b_i < 1$ and $\Sigma_\xi$ is also a diagonal matrix with diagonal elements $\sigma^2_{i,\xi} > 0$, for $i = 1, \ldots, N_z$; the other parts of the model are the same as in equations (1) - (2), but with the additional assumption that the disturbance $\xi_t$, for $t = 1, \ldots, T$, is mutually and serially uncorrelated with all other disturbances in the model. In our Monte Carlo study we set the parameters as $\Lambda_y = 1$, $\Phi = 0.8$, $\Sigma_\eta = 0.25$, $\lambda_{x,i} = 1/i$, $B = 0.8 \cdot I_{N_x}$, $\Sigma_\varepsilon = 0.5 \cdot I_{N_z}$, $\Sigma_\xi = 0.25 \cdot I_{N_z}$.

where $\lambda_{x,i}$ is the $(i,1)$ element of $\Lambda_x$, for $i = 1, \ldots, N_x$.

The second DGP for $z_t$ is the vector autoregressive process of order 1, the VAR(1) process

$$ z_t = \Psi z_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon), $$

(6)

where we assume that the autoregressive coefficient matrix $\Psi$ is an upper-triangular matrix that ensures a stationary process $z_t$ while variance matrix $\Sigma_\varepsilon$ is positive-definite. For the simulations, $\Sigma_\varepsilon = 0.5 \cdot I_{N_z}$, the diagonal elements of $\Psi$ are set to 0.8 while the upper diagonal elements are randomly chosen between $-0.5$ and 0.5 such that $z_t$ is stationary. This VAR(1) model ensures that the univariate time series $y_t$ is correlated with $x_{t-1}$ but the vector time series $x_t$ is not correlated with $y_{t-1}$.

In our simulation experiment, we consider two different settings. In the first case, we adopt the dynamic factor model as the DGP and we consider the same model for estimation and forecasting. We generate three sets of 500 vector time series $z_t$ with $N_y = 1$ and for $N_x = 2, 5$ and 10. For each simulated vector time series $z_t$, we estimate the parameters for the sequence of weight values $W = 1, \ldots, 5, 10, 50$ and 100. When the model is correctly specified, we expect that increasing the value of $W$ will not improve the forecasting accuracy for the variable of interest $y_t$. Theorem 2 has shown that asymptotically the different values of $W$ must yield the same results since the WMLE is consistent to the true parameter for any $W$. Any improvements in the correct specification setting are thus only finite-sample improvements. In the correct specification setting, we calculate the finite sample rejection rate of the Diebold-Mariano (DM) test under the null hypothesis $H_0$, which is the size of the DM test.

In the second case, we adopt the VAR(1) model (6) as DGP while the dynamic factor
model (5) is considered for the analysis. Similarly as in the first case, we generate 500 datasets from the VAR(1) process with \( N_x = 2, 5 \) and 10, and we estimate the parameters of the dynamic factor model (5) with values of \( W \) as stated above. In the misspecification case, we also consider the \( W \) values 250, 500 and 1000; here we expect that an increasing weight \( W \) will be beneficial for the forecasting accuracy of \( y_t \). Theorem 3 has shown that such large improvements are explained by the fact that we can use the weight \( W \) to let the estimated parameter vector converge to the pseudo-true parameter value. We also compute the finite-sample rejection rate of the DM test under the alternative hypothesis of not producing more accurate forecasts; this is the power of the DM test.

In Table 1, we present the mean squared error (MSE) averages for the variable of interest \( y_t \) for two cases. Each column is scaled with respect to the corresponding value for \( W = 1 \). From the right-hand side panel, we learn that in the misspecification case, increasing \( W \) leads to a better in-sample forecasting accuracy for \( y_t \), for all dimensions \( N_x \). However it is not necessary to choose very large values of \( W \). For example, for \( N_x = 2 \), the average value of MSE is smallest when we choose \( W = 250 \). Moreover, the improvements in the MSE appear to converge to some upper limit for \( N_x = 5, 10 \) when increasing the values for \( W \). Furthermore, we observe that more gains are made when more variables are included in the model such that the misspecification is more pronounced. On the other hand, for the correct specification case, the left-hand side panel of Table 1 reveal that the improvements of the in-sample forecasting accuracy are negligible for an increasing \( W \). For instance, the value of MSE is the smallest when \( W = 3 \) for \( N_x = 2 \), while the improvement is only about 0.04% to the benchmark. A large value of \( W \) does not necessarily lead to a better forecasting accuracy for \( y_t \). The forecasting accuracy of the WML method with \( W > 10 \) does not outperform the ML method.

In Table 2, we present the frequency of being chosen as the optimal weight for different value of \( W \) over 500 simulations. The main findings from Table 2 are similar to those from Table 1. Under correct specification, the weight \( W \) close to unity gives a more accurate in-sample forecasts. On the other hand, when the model is misspecified, the results suggest that we need to choose a large \( W \) in order to guarantee a better in-sample forecasting accuracy of \( y_t \).

In Table 3, we report the sample rejection rates of the DM test at 90% confidence level over 500 simulations for the forecasts obtained from WML parameter estimates against the forecasts obtained from ML parameter estimates. The left-hand side panel of Table 3 presents the results under correct model specification; this can be viewed as the size of the DM test in our setting. The right-hand side panel presents the results under misspecification this can be viewed as the power of the test. For \( N_x = 2 \), the rejection rate of the DM test is 11% and the test has roughly a correct size. For \( N_x = 5 \) and \( N_x = 10 \), the rejection rates are 14.4% and 12.2% so that we may conclude that the test is over-sized but not in a severe manner. On the other hand, the rejection rate of the DM test is 59.2% when \( N_x = 2 \) in the misspecification case and the rejection rate increases when we include more variables in \( x_t \). This indicates that the power of the test
is large when $N_x$ is large and small when $N_x$ is small. It supports the notion to work with large-scale dynamic factor models when accurate forecasts are required. We can conclude that for a larger dimension $N_x$, the DM test for the WML method is slightly oversized but the power of the DM test is strong.

<table>
<thead>
<tr>
<th>Correct Specification</th>
<th>Misspecification</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_x = 2$</td>
<td>$N_x = 5$</td>
</tr>
<tr>
<td>1</td>
<td>1.00000</td>
</tr>
<tr>
<td>2</td>
<td>0.99958</td>
</tr>
<tr>
<td>3</td>
<td>0.99958</td>
</tr>
<tr>
<td>4</td>
<td>0.99966</td>
</tr>
<tr>
<td>5</td>
<td>0.99999</td>
</tr>
<tr>
<td>10</td>
<td>1.00144</td>
</tr>
<tr>
<td>50</td>
<td>1.00753</td>
</tr>
<tr>
<td>100</td>
<td>1.01041</td>
</tr>
<tr>
<td>250</td>
<td>0.92192</td>
</tr>
<tr>
<td>500</td>
<td>0.92246</td>
</tr>
<tr>
<td>1000</td>
<td>0.92331</td>
</tr>
</tbody>
</table>

Table 1: Average MSE of target variable $y_t$ for different values of $W$. We present the average MSE for the target variable $y_t$, over 500 simulation runs. In the first case (correct specification) a DFM model with idiosyncratic AR(1) factors is used both for simulation as for estimation. In the second case (misspecification) the VAR(1) model is used for simulation while a DFM is used for estimation. The smallest MSE in each column is highlighted.

3 Mixed Frequency

Consider the case where we need to analyze variables which are observed at different frequencies by means of the dynamic factor model. More specifically, and most relevant for macroeconomic forecasting, we consider the case that the variable of interest $y_t$ is observed at a low frequency (quarterly) while the related variables in $x_t$ are typically observed at a high frequency (monthly, weekly, daily). The model (1) or the analysis based on this model needs to be modified in this case. We will discuss a number of solutions next.

3.1 Interpolation: MFI

The usual approach to a mixed frequency dynamic analysis is based on Mariano and Murasawa (2003). In terms of the dynamic factor model (1), although they consider the specification (5), the time index $t$ is considered to be a monthly (high frequency) index such that the model can accommodate the related variable $x_t$ in a standard fashion. The low frequency and key variable $y_t$ is also subject to the monthly index $t$ but for the months (or weeks, days, etc.) that no information is available for $y_t$, it is treated as a
Table 2: Frequencies for optimal weight values. This table presents the frequencies of the weights $W$ being chosen as optimal over 500 simulations. In the first case (correct specification) a DFM model with idiosyncratic AR(1) factors is used both for simulation as for estimation. In the second case (misspecification) the VAR(1) model is used for simulation while a DFM is used for estimation.

<table>
<thead>
<tr>
<th></th>
<th>Correct Specification</th>
<th>Misspecification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_x = 2$</td>
<td>$N_x = 5$</td>
</tr>
<tr>
<td>1</td>
<td>36.4</td>
<td>31.8</td>
</tr>
<tr>
<td>2</td>
<td>7.6</td>
<td>9.8</td>
</tr>
<tr>
<td>3</td>
<td>4.6</td>
<td>6.0</td>
</tr>
<tr>
<td>4</td>
<td>2.6</td>
<td>5.4</td>
</tr>
<tr>
<td>5</td>
<td>2.2</td>
<td>3.0</td>
</tr>
<tr>
<td>10</td>
<td>4.8</td>
<td>7.4</td>
</tr>
<tr>
<td>50</td>
<td>5.0</td>
<td>7.2</td>
</tr>
<tr>
<td>$\geq$100</td>
<td>16.2</td>
<td>9.4</td>
</tr>
</tbody>
</table>

Table 3: Realized Rejection Rate of DM test over 500 simulations. This table presents the finite sample rejection rates at 90% confidence level of the DM test over 500 simulations. The results in the left panel can be considered as the size of the test and the right panel as the power of the test. In the first case (correct specification) a DFM model with idiosyncratic AR(1) factors is used both for simulation as for estimation. In the second case (misspecification) the VAR(1) model is used for simulation while a DFM is used for estimation.

<table>
<thead>
<tr>
<th></th>
<th>Correct Specification</th>
<th>Misspecification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_x = 2$</td>
<td>$N_x = 5$</td>
</tr>
<tr>
<td>Realized Rejection Rate</td>
<td>0.110</td>
<td>0.144</td>
</tr>
</tbody>
</table>
missing value. Parameter estimation and forecasting is operated by the Kalman filter that can handle missing values in an appropriate manner. An important application of this approach for the nowcasting and forecasting of GDP growth in the euro area is carried out by Bărbura, Giannone, Modugno, and Reichlin (2013) on the basis of a high-dimensional dynamic factor model.

3.2 Averaging: MFA

An alternative approach, and a natural counterpart to the MFI method above, is to have the time index $t$ as a quarterly index. In this case the variable of interest $y_t$ can be treated in a standard manner by the dynamic factor model (1). To incorporate the high frequency (monthly) related variables in this analysis, we can simply average their observations into the low frequency variable $x_t$. In other words, we have $x_t$ as the three-monthly (or 13-weekly) average of the variable, within each quarter $t$. The analysis is then simply reduced to a quarterly analysis based on (1). We should emphasize that nowcasts and monthly forecasts cannot be generated by this dynamic factor model.

3.3 Stacking high frequency variables into low frequency vector

The MFI method has the advantage that it preserves the dynamic features of the quarterly and monthly variables in the analysis while the MFA method disregards the monthly dynamics completely. The MFI method is a somewhat artificial solution as it requires the imputation of a long periodic sequence of missing values for the key variable $y_t$. Furthermore, in many cases it does not lead to a computational efficient solution. In Appendix B we argue that monthly dynamics can also effectively be treated by a model formulated in terms of a low frequency time index, say quarterly index. It is argued that any monthly linear dynamic process can be formulated by a multivariate model with a quarterly time index through the stacking of monthly observations in a quarterly vector time series. These ideas of stacking the series observed at higher frequencies into vectors of the lowest frequency have also been used by Ghysels (2012) in a vector autoregressive context and by Marcellino, Carriero, and Clark (2014) in the right-hand side of their Bayesian regression model with stochastic volatility. In our case we adopt these ideas in the left-hand side of a dynamic factor model. More references are given in the introductory section and more details are provided in Appendix B.

Let the monthly (high frequency) index be denoted by $\tau$ and the quarterly (low frequency) index by $t$. Consider an univariate monthly observed time series denoted by $x^m_\tau$ which we model by the monthly autoregressive process of order 1, denoted by AR(1), given by $x^m_\tau = \phi_x x^m_{\tau-1} + \varepsilon^m_\tau$ with the autoregressive parameter $\phi_x$ and Gaussian disturbance $\varepsilon^m_\tau \sim N(0, \sigma^2_{\varepsilon})$. We further have a quarterly observed times series $y_t$ which is modeled by the quarterly AR(1) process, given by $y_t = \phi_y y_{t-1} + \xi_t$, with $\xi_t \sim N(0, \sigma^2_{\xi})$, where $\phi_y$ is the autoregressive coefficient for the quarterly lagged dependent variable $y_{t-1}$ and $\xi_t$ is the Gaussian disturbance that is possibly correlated with $\varepsilon^m_\tau$. Further details on notation
and on low frequency model formulations for high frequency autoregressive processes, see Appendix B.

We stack three consecutive values of the monthly variable $x_m^\tau$ corresponding to a specific quarter $t$ into the quarterly $3 \times 1$ vector $x_t$, as in Appendix B, that is

$$x_t = \begin{pmatrix} x_{t,1} \\ x_{t,2} \\ x_{t,3} \end{pmatrix} = \begin{pmatrix} x^m_\tau \\ x^m_{\tau+1} \\ x^m_{\tau+2} \end{pmatrix}. \tag{7}$$

Then the two processes for $y_t$ and $x_t$ can be combined into the low frequency vector process

$$\begin{pmatrix} y_t \\ x_{t,1} \\ x_{t,2} \\ x_{t,3} \end{pmatrix} = \begin{pmatrix} \phi_y & 0 & 0 & 0 \\ 0 & 0 & 0 & \phi_x \\ 0 & 0 & 0 & \phi_x^2 \\ 0 & 0 & 0 & \phi_x^3 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1,1} \\ x_{t-1,2} \\ x_{t-1,3} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \phi_x & 1 & 0 \\ 0 & \phi_x^2 & \phi_x & 1 \end{pmatrix} \begin{pmatrix} \xi_t \\ \varepsilon_{t,1} \\ \varepsilon_{t,2} \\ \varepsilon_{t,3} \end{pmatrix}, \tag{8}$$

for $t = 1, \ldots, T$. Appendix B provides the derivation of this formulation. It is shown that the distributional properties of the monthly series $x^m_\tau$ and the quarterly series $x_t$ are the same. In particular, the quarterly vector process is just a different formulation of the monthly process but the dynamic properties are not altered. We further notice the difference between the autoregressive parameters $\phi_x$ and $\phi_y$. The parameter $\phi_x$ measures the monthly dependence of $x^m_\tau$ on its lagged value $x^m_{\tau-1}$ of one month earlier, whereas the parameter $\phi_y$ indicates the dependence of $y_t$ on its lagged value $y_{t-1}$ of one quarter earlier.

In model (8) the two processes for $y_t$ and $x_t$ are independent of each other, but we can allow for covariances between $\xi_t$ and elements of $\varepsilon_t$. Also, the linear dependence between $y_t$ and $x_{t-1}$ can be introduced straightforwardly in this model. The direct dependence between $y_t$ and $x_t$ can also be established in the dynamic factor model by introducing a pre-multiplication matrix on the left-hand side of (1).

### 3.4 Computing Times

In Appendix B we show how the mixed frequency solution of (8) can be generalized to any AR($p$). Whether the monthly AR model is represented by the monthly process $x^m_\tau$ or by the stacked quarterly $3 \times 1$ vector $x_t$, or a yearly $12 \times 1$ vector, it has no effect on the value of the loglikelihood function for a given parameter vector. The low and high frequency representations are observationally equivalent as the derivations only rely on equalities; see Appendix B. In all cases, the Kalman filter can be used for likelihood evaluation. Hence the maximized loglikelihood value and the corresponding parameter estimates are the same for low and high frequency representations.

However, the representation has an effect on computing times. For example, for 100 years of data and for a monthly representation, we have $T = 1200$. When the
data is stacked into quarterly $3 \times 1$ vectors we have $T = 400$ and with yearly $12 \times 1$ vectors we only have $T = 100$. On the other hand, the stacked vectors become larger when higher frequency processes are represented by a low frequency model. Hence the different representations will have an effect on the computation times when evaluating the loglikelihood function.

To illustrate this, we have evaluated the loglikelihood value 10,000 times for simulated AR($p$) models of a length of 1,000 years, for different orders $p$ and using different representations: daily, weekly, monthly, quarterly and yearly. For example, for a monthly AR($p$) process, time series are generated consisting of $T = 12,000$ monthly observations. The loglikelihood value was then calculated 10,000 times using the parameter values that maximized the loglikelihood function. We have verified that all likelihood evaluations resulted in the same value.

The computing times for different combinations of AR($p$) processes and frequencies are presented in Table 3.4. We focus here on the results for weekly and daily autoregressive processes; we present results for monthly processes in Appendix B. It is clear that for weekly and daily AR(1) and AR(2) processes, the representations based on weekly and daily models, respectively, are most efficient. For these cases a lower-dimensional stacked vector outweighs the fact that the Kalman filter has to go through 52,000 (weekly) and 364,000 (daily) iterations instead of 1,000 iterations in their respective yearly representations. But for a weekly AR(3) process, or for any order $p > 2$, the 13-monthly representation (we assume that each month consists of 4 weeks) leads to a faster computation of the likelihood function. Here the smaller time dimension is beneficial while the size of the stacked vectors are of the same size given the number of lags that need to be accommodated. Similar effects take place when $p$ increases further and even the yearly representation become the computationally more efficient one. The results for weekly and daily processes in Table 3.4 clearly illustrate that stacking observations into lower frequency vectors can lead to large computational gains, especially when many lagged dependent variables are part of the model.

### 3.5 Stacking for Dynamic Factor Model, monthly case MFS-M

We discuss how the dynamic factor model needs to be modified when the solution of stacking for mixed frequencies is considered. Consider we have a large vector of $x^m_\tau$ of high frequency variables that are potentially useful for the forecasting of the low frequency variable of interest $y_t$. We modify the dynamic factor (1) for a quarterly time index $t$ but with $r \times 1$ vector with dynamic factors $f^m_\tau$, for a specific month $\tau$, that is subject to a monthly dynamic process. For example, we can consider the vector autoregressive process of order $p^m$ given by

$$f^m_\tau = \Phi^m_{f,1} f^m_{\tau-1} + \Phi^m_{f,2} f^m_{\tau-2} + \ldots + \Phi^m_{f,p^m} f^m_{\tau-p^m} + \eta^m_\tau, \quad \eta^m_\tau \sim N(0, \Sigma_\eta), \tag{9}$$
Table 4: Computing times. The left panel of this table presents the total computing time that is required to filter 1,000 weekly time series with $T = 52,000$ using Kalman filter for the corresponding AR($p$) model. Four different approaches are used: treating the data as weekly observations, stacking the data into 13 "monthly" $4 \times 1$ vectors, stacking the data into quarterly $13 \times 1$ vectors and stacking the data into yearly $52 \times 1$ vectors. The right panel of this table presents the total computing time that is required to filter 1,000 daily time series with $T = 364,000$ using Kalman filter for the corresponding AR($p$) model. Four different approaches are used: treating the data as daily observations, stacking the data into weekly $7 \times 1$ vectors, stacking the data into 13 "monthly" $28 \times 1$ vectors and stacking the data into quarterly $91 \times 1$ vectors. Each value presents the aggregate computing time over 1000 simulations. For each $p$ the fastest of the four approaches is highlighted.

<table>
<thead>
<tr>
<th>$p$</th>
<th>Week</th>
<th>&quot;Month&quot;</th>
<th>Quarter</th>
<th>Year</th>
<th>$p$</th>
<th>Day</th>
<th>Week</th>
<th>&quot;Month&quot;</th>
<th>Quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.5</td>
<td>3.5</td>
<td>6.9</td>
<td>69.9</td>
<td>1</td>
<td>2.6</td>
<td>3.9</td>
<td>15.3</td>
<td>108.4</td>
</tr>
<tr>
<td>2</td>
<td>3.4</td>
<td>3.8</td>
<td>7.5</td>
<td>72.5</td>
<td>2</td>
<td>3.6</td>
<td>4.2</td>
<td>16.6</td>
<td>112.7</td>
</tr>
<tr>
<td>3</td>
<td>4.4</td>
<td>4.0</td>
<td>8.3</td>
<td>76.1</td>
<td>3</td>
<td>4.6</td>
<td>4.5</td>
<td>17.7</td>
<td>116.4</td>
</tr>
<tr>
<td>6</td>
<td>14.0</td>
<td>7.8</td>
<td>10.6</td>
<td>85.4</td>
<td>6</td>
<td>14.6</td>
<td>7.0</td>
<td>21.8</td>
<td>130.7</td>
</tr>
<tr>
<td>7</td>
<td>18.1</td>
<td>10.1</td>
<td>11.5</td>
<td>87.3</td>
<td>7</td>
<td>18.8</td>
<td>9.2</td>
<td>29.7</td>
<td>132.0</td>
</tr>
<tr>
<td>8</td>
<td>24.3</td>
<td>11.9</td>
<td>12.2</td>
<td>89.9</td>
<td>8</td>
<td>24.9</td>
<td>11.5</td>
<td>24.4</td>
<td>135.3</td>
</tr>
<tr>
<td>11</td>
<td>47.0</td>
<td>22.4</td>
<td>14.4</td>
<td>100.2</td>
<td>18</td>
<td>379.4</td>
<td>57.2</td>
<td>36.6</td>
<td>174.5</td>
</tr>
<tr>
<td>12</td>
<td>54.2</td>
<td>24.9</td>
<td>15.2</td>
<td>103.2</td>
<td>19</td>
<td>287.3</td>
<td>68.5</td>
<td>39.9</td>
<td>179.0</td>
</tr>
<tr>
<td>13</td>
<td>65.6</td>
<td>30.1</td>
<td>18.2</td>
<td>106.0</td>
<td>20</td>
<td>319.8</td>
<td>68.3</td>
<td>38.8</td>
<td>182.2</td>
</tr>
<tr>
<td>46</td>
<td>6504.6</td>
<td>609.5</td>
<td>325.3</td>
<td>209.9</td>
<td>58</td>
<td>13333.8</td>
<td>930.8</td>
<td>435.4</td>
<td>316.3</td>
</tr>
<tr>
<td>47</td>
<td>8064.1</td>
<td>647.5</td>
<td>354.2</td>
<td>209.3</td>
<td>59</td>
<td>16124.3</td>
<td>970.4</td>
<td>455.5</td>
<td>318.6</td>
</tr>
<tr>
<td>48</td>
<td>9252.8</td>
<td>684.8</td>
<td>363.3</td>
<td>220.5</td>
<td>60</td>
<td>15937.9</td>
<td>1005.2</td>
<td>470.9</td>
<td>323.1</td>
</tr>
</tbody>
</table>
where $\Phi_{f_{j}}$ is the autoregressive coefficient matrix for lag $j$ and for the monthly vector $f_{m}$ and $\eta_{m}$ is the disturbance vector. The stack of three consecutive months corresponding to a specific quarter $t$ is $x_{t}$ and is defined in (7). In a similar way, we can define the stacked vector of dynamic factors, $f_{t} = (f_{t1}, f_{t2}, f_{t3})'$ = $(f_{m}^{m}, f_{m}^{m+1}, f_{m}^{m+2})'$ and the stacked disturbance vector $\eta_{t}$. We obtain the mixed frequency dynamic factor model

$$
\begin{pmatrix}
y_t \\
x_{t,1} \\
x_{t,2} \\
x_{t,3}
\end{pmatrix} =
\begin{pmatrix}
\Lambda_y & \Lambda_y & \Lambda_y \\
\Lambda_x & 0 & 0 \\
0 & \Lambda_x & 0 \\
0 & 0 & \Lambda_x
\end{pmatrix}
\begin{pmatrix}
f_{t,1} \\
f_{t,2} \\
f_{t,3}
\end{pmatrix} +
\begin{pmatrix}
e_{y,t} \\
e_{x,t,1} \\
e_{x,t,2} \\
e_{x,t,3}
\end{pmatrix}
$$

(10)

where factor loading matrix $\Lambda = [\Lambda'_y, \Lambda'_x]'$ is effectively the same as in (1). The similarity with (1) becomes even more pronounced by formulating the model as

$$
\begin{pmatrix}
y_t \\
x_t
\end{pmatrix} =
\begin{pmatrix}
\iota_3 \otimes \Lambda_y \\
I_3 \otimes \Lambda_x
\end{pmatrix} f_t +
\begin{pmatrix}
e_{y,t} \\
e_{x,t}
\end{pmatrix},
$$

where $\iota_k$ is the $k \times 1$ vector of ones. The dynamic specification for $f_t$, based on the autoregressive process or the linear dynamic process for the monthly vector $f_{m}$, can be formulated in state space form as is discussed in Appendix B. In particular, we can write

$$
f_t = Z_f \cdot \alpha_t, \quad \alpha_t = T_f \alpha_t + R_f \zeta_t, \quad \zeta_t \sim N(0, \Sigma_\zeta),
$$

(11)

where $Z_f$ is a selection matrix, $\alpha_t$ is the state vector, $T_f$ is the transition matrix and where the variance structure of the state vector is determined by $R_f$ and $\Sigma_\zeta$. More general specifications can also be considered. For example, we can replace $\iota_3 \otimes \Lambda_y$ by a matrix with three different loading matrices for each month, that is $\Lambda_y = [\Lambda_{y1}, \Lambda_{y2}, \Lambda_{y3}]$. Other mixed frequencies than for monthly and quarterly variables can be considered. For example, a mix of yearly and weekly variables can be jointly modelled using the above stacking approach in a similar way. Furthermore, the generality of the state space framework is discussed in Appendix B and can be fully exploited.

### 3.6 Stacking for Dynamic Factor Model, quarterly case MFS-Q

We can also opt for a quarterly dynamic factor directly, rather than the monthly factor $f_{m}$ as discussed above. In this case, the vector $f_{t}$ is not a stacked vector; vector $f_{t}$ only contains the $r$ factors which are specified as quarterly dynamic processes. For example, we can specify dynamic process for $f_{t}$ by the vector autoregressive model of order $p$, VAR($p$),

$$
f_t = \Phi_1 f_{t-1} + \ldots + \Phi_p f_{t-p} + \eta_t, \quad \eta_t \sim N(0, \Sigma_\eta),
$$

with autoregressive coefficient matrix $\Phi_j$, for lag $j = 1, \ldots, p$ in quarters. The VAR($p$) model can also be formulated in the state space form (11). The observation equation of
the dynamic factor model for quarterly factors is then simply given by

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{bmatrix} \Lambda_y \\ \tau_3 \otimes \Lambda_x \end{bmatrix} f_t + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix},$$

(12)

In this case we only have one value for $f_t$ for each quarter, for all three months.

4 Empirical Study: forecasting U.S. GDP growth

4.1 Design of study

In our empirical study, we investigate the forecasting performances of our proposed models with their parameters estimated by weighted maximum likelihood. We focus on the forecasting of U.S. GDP growth and adopt the standard database of Stock and Watson (2005) that consists of 132 macroeconomic variables. This will facilitate the comparisons with other empirical studies. We consider four different dynamic factor model (DFM) specifications together with one benchmark model. The four DFMs differ in their treatment of mixed frequency data. Our first model (mixed frequency interpolation, MFI) is based on the model of Mariano and Murasawa (2003) which is formulated in a monthly frequency. The key variable GDP growth has a quarterly frequency and is incorporated in the model by having missing values for the months that the new GDP value is not yet available. The remaining three models are formulated in the quarterly frequency. The second model (average, MFA) simply treats all variables as quarterly variables; three consecutive monthly observations are averaged to obtain the quarterly value. The third model (stacked with monthly factors, MFS-M) treats monthly variables as a three-dimensional vector of quarterly observations, preserves their monthly dynamics, and has the dynamic factors as monthly variables. The fourth model (stacked with quarterly factors, MFS-Q) is the same as the MFS-M model but has the dynamic factors as quarterly variables. The details of the last three models are described in Section 3. All these mixed frequency dynamic factor models are extended with idiosyncratic autoregressive components for each variable in $z_t = (y_t', x_t')'$ as in the model specification of Mariano and Murasawa (2003). The benchmark model is the well-known Bridge model of Trehan (1989). We assess the improvements in forecasting and nowcasting accuracy by adopting the method of weighted maximum likelihood for different weights $W$; see the discussion in Section 2. We evaluate the forecasting and nowcasting accuracy for the various methods and compare the results, in terms of mean squared error (MSE) and the Diebold-Mariano (DM) test.

Our dynamic factor modeling framework is similar to the one proposed by Bräuning and Koopman (2014) where the high-dimensional vector of related variables is collapsed to a smaller set of principal components. In particular, we have $N_y = 1$ for the variable of interest, GDP growth, and $N_x = 7$ for the monthly principal components obtained from the Stock and Watson (2005) data set. The seven principal components are computed
as the first principal component from seven different groups of variables in the data set. The details are provided next.

4.2 Data

We adopt the dataset of Stock and Watson (2005) for the forecasting of quarterly U.S. GDP. This data set includes 132 real economic indicators, which are observed at a monthly frequency from January 1960 to December 2009. The out-of-sample forecasting period starts from January 2000 and ends at December 2009. We construct seven principal components that correspond to the highest eigenvalue of the sample covariance matrix of variables from seven different subsets of the data set. Each subset correspond to a category of economic indicators; see the list of categories in Table 5. All the data are transformed and demeaned so that no intercept coefficients are required in the model. Detected outliers in each series are replaced by their median value of the previous five observations; here we follow Stock and Watson (2005).

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Description</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP</td>
<td>U.S. Real GDP (billions of chained 1996)</td>
<td>Q</td>
</tr>
<tr>
<td>OUT</td>
<td>First PC from category &quot;Real output and income&quot;, 18 variables</td>
<td>M</td>
</tr>
<tr>
<td>EMP</td>
<td>First PC from &quot;Employment and hours&quot;, 30 variables</td>
<td>M</td>
</tr>
<tr>
<td>RTS</td>
<td>First PC from &quot;Real retail, manufacturing and trade sales&quot;, 2 variables</td>
<td>M</td>
</tr>
<tr>
<td>PCE</td>
<td>Consumption, 1 variable</td>
<td>M</td>
</tr>
<tr>
<td>HHS</td>
<td>First PC from &quot;Housing starts and sales&quot;, 10 variables</td>
<td>M</td>
</tr>
<tr>
<td>ORD</td>
<td>First PC from &quot;Inventories and Orders&quot;, 10 variables</td>
<td>M</td>
</tr>
<tr>
<td>OTH</td>
<td>First PC from category &quot;Other&quot;, including stock prices, exchange rates, interest rates, money and credit, etc., 61 variables</td>
<td>M</td>
</tr>
</tbody>
</table>

Table 5: Data definitions. This table presents the definitions of all the quarterly, monthly and semi-monthly variables that are used in our empirical study. Entries in the third column present the frequency of the series: monthly(M) and quarterly(Q). The principal component is referred to as PC.

4.3 Empirical results

We compare the forecast accuracies based on the mean square error (MSE). The sample period from quarter 1, 2000 to quarter 4, 2009 (2000Q1-2009Q4) is used to evaluate the forecasting performance. The forecasts are calculated using a rolling window of 20 years of data to estimate the parameters. We evaluate both the nowcasting \((h = 1, 2)\) and the forecasting \((h = 3, 6, 9, 12)\) performance of all the competing models. The forecasts are calculated at a monthly frequency. When \(h = 1\), the values of \(x_t\) are known until the first two months of the quarter that is to be forecasted. When \(h = 2\), only the first month of the quarter that we want to forecast is observed. When \(h = 3\), we are forecasting one quarter ahead, no observations are available for the quarter that we forecast. All values
until the previous quarter are observed. Similarly, when \( h = 6, 9, 12 \) we are forecasting two, three and four quarters ahead respectively. In the MFA model, there are no monthly dynamics, so it is only possible to forecast at \( h = 3, 6, 9, 12 \). The single common factor \( f_t \) is specified as an AR(1) process while all idiosyncratic factors are taken as AR(2) processes for all models. Finally, for each model, the factor loading for the quarterly GDP growth equation is set to the unity value for identification purposes.

We also evaluate the forecasting performance of MSF-M and MSF-Q using the weighted maximum likelihood (WML) method. A cross validation analysis is implemented as follows. First, the WML parameter, \( W \), is estimated using forecasting period 1980Q1-1999Q4. The in-sample results are based on this estimation sample. For the estimated \( W \), we compute the forecasts of the period 2000Q1-2009Q4 and the out-of-sample forecasts can be evaluated for all time periods.

<table>
<thead>
<tr>
<th></th>
<th>( h = 1 )</th>
<th>( h = 2 )</th>
<th>( h = 3 )</th>
<th>( h = 6 )</th>
<th>( h = 9 )</th>
<th>( h = 12 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ML</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>(1.000)</td>
<td>(1.000)</td>
<td>(1.000)</td>
<td>(1.000)</td>
<td>(1.000)</td>
<td>(1.000)</td>
</tr>
<tr>
<td>MFI</td>
<td>1.065</td>
<td>1.136</td>
<td>1.044</td>
<td>1.069</td>
<td>1.040</td>
<td>1.013</td>
</tr>
<tr>
<td></td>
<td>(0.672)</td>
<td>(0.915)</td>
<td>(0.692)</td>
<td>(0.890)</td>
<td>(0.761)</td>
<td>(0.650)</td>
</tr>
<tr>
<td>MFS-M</td>
<td>1.221</td>
<td>1.152</td>
<td>0.968</td>
<td>0.992</td>
<td>1.029</td>
<td>1.001</td>
</tr>
<tr>
<td></td>
<td>(0.975)</td>
<td>(0.942)</td>
<td>(0.210)</td>
<td>(0.401)</td>
<td>(0.856)</td>
<td>(0.518)</td>
</tr>
<tr>
<td>MFS-Q</td>
<td>1.022</td>
<td>0.973</td>
<td>0.916</td>
<td>0.993</td>
<td>1.070</td>
<td>1.040</td>
</tr>
<tr>
<td></td>
<td>(0.607)</td>
<td>(0.380)</td>
<td>(0.147)</td>
<td>(0.455)</td>
<td>(0.877)</td>
<td>(0.749)</td>
</tr>
<tr>
<td>MFA</td>
<td>0.981</td>
<td>0.970</td>
<td>1.004</td>
<td>0.990</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.317)</td>
<td>(0.142)</td>
<td>(0.549)</td>
<td>(0.328)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>WML</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFS-M</td>
<td>1.221</td>
<td>1.152</td>
<td>0.968</td>
<td>0.992</td>
<td>1.029</td>
<td>1.001</td>
</tr>
<tr>
<td></td>
<td>(0.975)</td>
<td>(0.942)</td>
<td>(0.210)</td>
<td>(0.401)</td>
<td>(0.856)</td>
<td>(0.518)</td>
</tr>
<tr>
<td>MFS-Q</td>
<td>0.984</td>
<td>0.943</td>
<td>0.918*</td>
<td>0.993</td>
<td>0.994</td>
<td>1.015</td>
</tr>
<tr>
<td></td>
<td>(0.434)</td>
<td>(0.294)</td>
<td>(0.080)</td>
<td>(0.439)</td>
<td>(0.426)</td>
<td>(0.719)</td>
</tr>
</tbody>
</table>

Table 6: Forecasting comparisons for the quarterly US DGP growth rate from 2000Q1 till 2009Q4 at forecasting horizon \( h = 1, 2, 3, 6, 9, 12 \). We present the forecast RMSE ratios of the competing models relative to the benchmark Bridge model (BM). We consider four dynamic factor models as the competing models, which are mixed frequency Interpolation (MFI), mixed frequency stacking with monthly factors (MFS-M), mixed frequency stacking with quarterly factors (MFS-Q), and mixed frequency aggregation (MFA). All results are based on parameter estimates obtained form the 20-year rolling window starting from 1980Q1. For each forecasting horizon the most accurate model is highlighted.

The forecasting results for the four mixed frequency dynamic factor models are compared with the benchmark model in the first panel of Table 6. The MFS-Q model provides the most accurate forecasts for the GDP growth rate among the dynamic factor models when nowcasting and at the forecasting horizon \( h = 3 \). The accuracy is higher than the original MFI model for all forecasting horizons. The quarterly MFA model is more accurate for forecasts made at longer horizons, for \( h = 6, 9, 12 \). Furthermore, the MFS-Q
model is more accurate than the original MFI model for these forecast horizons. When we further take into account the benchmark Bridge model, we find that the forecasts of the Bridge model are most accurate at \( h = 0 \) and \( h = 9 \).

We adopt the Diebold-Mariano (DM) test to verify whether forecast accuracy of dynamic factor models are significantly different from the benchmark. The \( p \) values of the DM test are also presented in Table 6. Although the MSE ratios for the dynamic factor models are smaller than unity at the different forecasting horizons, none of these models are significantly better than the benchmark, for any horizon and at the 10% level. For example, MFS-M model has a smaller MSE than the benchmark model at horizon \( h = 3, 6 \) but the \( p \) values of the DM test has larger values than 0.1. The MFS-Q model has a smaller MSE than the benchmark at horizon \( h = 2, 3, 6 \) but the \( p \) values of the DM test are larger; only at \( h = 3 \) the \( p \) value is close to 0.1.

<table>
<thead>
<tr>
<th></th>
<th>( h = 1 )</th>
<th>( h = 2 )</th>
<th>( h = 3 )</th>
<th>( h = 6 )</th>
<th>( h = 9 )</th>
<th>( h = 12 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFS-Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>2</td>
<td>0.81878</td>
<td>0.82198</td>
<td>0.91035</td>
<td>0.94619</td>
<td>0.93689</td>
<td>0.94538</td>
</tr>
<tr>
<td>3</td>
<td>0.83601</td>
<td>0.79331</td>
<td>0.88796</td>
<td>0.93955</td>
<td>0.94181</td>
<td>0.93350</td>
</tr>
<tr>
<td>4</td>
<td>0.81675</td>
<td>0.78928</td>
<td>0.88118</td>
<td>0.93150</td>
<td>0.92853</td>
<td>0.93555</td>
</tr>
<tr>
<td>5</td>
<td>0.83421</td>
<td>0.78999</td>
<td>0.88349</td>
<td>0.93410</td>
<td>0.93216</td>
<td>0.93604</td>
</tr>
<tr>
<td>6</td>
<td>0.83999</td>
<td>0.78485</td>
<td>0.87975</td>
<td>0.93356</td>
<td>0.94094</td>
<td>0.93754</td>
</tr>
<tr>
<td>7</td>
<td>0.84836</td>
<td>0.78396</td>
<td>0.87840</td>
<td>0.94051</td>
<td>0.94526</td>
<td>0.93420</td>
</tr>
<tr>
<td>8</td>
<td>0.85676</td>
<td>0.78676</td>
<td>0.87027</td>
<td>0.94227</td>
<td>0.94674</td>
<td>0.93618</td>
</tr>
<tr>
<td>9</td>
<td>0.86989</td>
<td>0.79076</td>
<td>0.86854</td>
<td>0.94492</td>
<td>0.93127</td>
<td>0.94295</td>
</tr>
<tr>
<td>10</td>
<td>0.87273</td>
<td>0.79033</td>
<td>0.87766</td>
<td>0.94750</td>
<td>0.94862</td>
<td>0.94201</td>
</tr>
<tr>
<td>MFS-M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>2</td>
<td>1.01600</td>
<td>1.01600</td>
<td>1.01600</td>
<td>1.01381</td>
<td>1.00225</td>
<td>0.99964</td>
</tr>
<tr>
<td>3</td>
<td>0.85002</td>
<td>0.83158</td>
<td>0.91551</td>
<td>1.00590</td>
<td>0.99455</td>
<td>1.01751</td>
</tr>
<tr>
<td>4</td>
<td>0.71480</td>
<td>0.73613</td>
<td>0.88738</td>
<td>1.00750</td>
<td>0.99112</td>
<td>1.03126</td>
</tr>
<tr>
<td>5</td>
<td>0.69002</td>
<td>0.70639</td>
<td>0.85165</td>
<td>0.96897</td>
<td>0.95040</td>
<td>0.99944</td>
</tr>
<tr>
<td>6</td>
<td>0.68837</td>
<td>0.70099</td>
<td>0.85427</td>
<td>0.98279</td>
<td>0.97446</td>
<td>1.00560</td>
</tr>
<tr>
<td>7</td>
<td>0.69169</td>
<td>0.70450</td>
<td>0.85604</td>
<td>0.98496</td>
<td>0.96951</td>
<td>1.02160</td>
</tr>
<tr>
<td>8</td>
<td>0.71637</td>
<td>0.73022</td>
<td>0.86932</td>
<td>0.98690</td>
<td>0.98272</td>
<td>1.01616</td>
</tr>
<tr>
<td>9</td>
<td>0.71382</td>
<td>0.71678</td>
<td>0.84706</td>
<td>0.99015</td>
<td>0.97717</td>
<td>1.02232</td>
</tr>
<tr>
<td>10</td>
<td>0.70234</td>
<td>0.70515</td>
<td>0.85230</td>
<td>0.98552</td>
<td>0.98456</td>
<td>1.02290</td>
</tr>
</tbody>
</table>

Table 7: Forecasting comparisons for US GDP growth using MFS-Q and MFS-M, based on ML and weighted ML (WML). The RMSE ratios of the forecasts are presented for the quarterly observed US GDP growth based on the mixed frequency dynamic factor model, at forecasting horizons \( h = 1, 2, 3, 6, 9, 12 \) using different \( W \) values in the WML procedure. The data sample period is from 1960Q1 till 1999Q4. For each forecasting horizon the most accurate model is highlighted.

In the second panel of Table 6, we present the results for the weighted maximum likelihood (WML) method applied to the MFS-M and MFS-Q models. We take different integer values for \( W \) and determine its optimal value in the estimation sample. We adopt
effectively an cross validation approach. The value of $W$ is selected using the sample from 1980Q1 until 1999Q4. The optimal value of $W$ is adopted for the forecasting exercise; see Table 7 for the MFS-Q and MFS-M models. The optimal weight at each forecasting horizon is highlighted. We notice that the weight $W$ is not limited to integer values only. We have also investigated the MSE of the WML method for the out-of-sample period. Some values of $W$ around the optimal value found in Table 7 can still yield a smaller MSE but the improvement is negligible and far from significant. Hence these initial results already provide a clear picture of the usefulness of the WML method.

The out-of-sample forecast MSEs, for the period from 2000Q1 to 2009Q4, and based on WML are presented in Table 6. They can be compared with the forecasting results from the other dynamic factor and benchmark models. We may conclude that the WML method improves the forecasting accuracy in the cross-validation sample but also provides accurate out-of-sample forecasts in various cases. For example, the WML method leads to a 3.7% improvement in MSE at $h = 1$, a 3.1% improvement at $h = 2$ and a 7.2% improvement in MSE at $h = 9$ for MFS-Q model. For MFS-Q model at $h = 3$, the standard ML method does not lead to a smaller MSE, but when applying WML method, the MFS-Q model is significantly better than the benchmark at 10% confidence level.

We have shown that our stacking approach can treat mixed frequency data sets in a standard fashion. The WML method is able to produce more accurate forecasts for shorter forecast horizons.

5 Conclusions

This paper has introduced a novel weighted maximum likelihood (WML) estimation procedure for a class of mixed frequency dynamic factor models. The WML introduces a variable-specific weight in the likelihood function to let some variable equations be of more importance during the estimation process. We have derived the asymptotic properties of the weighted ML estimator and have provided an information-theoretic characterization of the WML estimator based on the Kullback-Leibler divergence. Furthermore, we have proposed a cross-validation method to estimate the weights that optimize the forecasting performance of the model and an $F$-statistic to test the statistical significance of the improvements in forecasting accuracy. A Monte Carlo study is carried out to investigate the finite-sample behavior of the WML estimator; it highlights its good performance. Furthermore we have shown that the low frequency representation of high frequency variables in a dynamic factor model provides a computational efficient treatment of mixed frequency data. The representation is flexible and can allow for low and high frequency factors. The presented illustration shows that our solutions lead to empirically relevant improvements in nowcasting and forecasting of GDP growth. We expect that our proposed solutions also have consequences in other applications and in other modeling frameworks.
References


Technical Appendix

Weighted Maximum Likelihood Estimator for Mixed Frequency Dynamic Factor Models

Francisco Blasques, Siem Jan Koopman, Max Mallee and Zhaokun Zhang
VU University Amsterdam, Department of Econometrics
Tinbergen Institute Amsterdam

A Proofs of Theorems and Propositions

Proof of Proposition 1. Let \( \{f_t(f_1^*)\}_{t \in \mathbb{N}} \) be generated according to (9) with initialization \( f_1^* \) and \( \|T_f\| < 1, \|R_f\| < \infty \) and \( \|\Sigma_2^g\| < \infty \). Then by Theorem 3.1 in Bougerol (1993), \( \{f_t(f_1^*)\}_{t \in \mathbb{N}} \) converges to an SE sequence \( \{f_t\}_{t \in \mathbb{Z}} \) satisfying \( \mathbb{E}|f_t|^{r} < \infty \forall r \). Uniqueness of the limit SE sequence is obtained in Straumann and Mikosch (2006). Furthermore, since \( \{f_t\} \) is a linear Gaussian process with \( \mathbb{E}|x_t|^{r} < \infty \forall r > 0 \). The bounds \( |\beta_x| < \infty \) and \( 0 < \sigma_x^2 < \infty \), together with the iid Gaussian nature of the innovations \( \{\epsilon_t\} \) ensure that \( \{x_t\} \) is SE and Gaussian with bounded moments of any order. Similarly, the bounds \( |\beta_y| < \infty \) and \( 0 < \sigma_y^2 < \infty \) and the iid Gaussian nature of \( \{\zeta_t\} \) ensure the SE linear Gaussian nature of \( \{y_t\} \) with \( \mathbb{E}|y_t|^{r} < \infty \forall r > 0 \).

Proof of Theorem 1. For every given \( w \in [0, 1] \), the random likelihood function \( L_T(\cdot, \hat{f}_1^*) \) is trivially almost surely continuous on \( \Psi \). The compactness of \( \Psi \) implies by Weierstrass’ theorem that the arg max set is almost surely non-empty. As a result, \( \hat{\psi}_T \) exists almost surely \( \forall T \in \mathbb{N} \). The continuity of the likelihood function in \( f_t, x_t \) and \( y_t \) for every \( \psi \in \Psi \) implies also measurability of the likelihood under the Borel \( \sigma \)-algebra. For every given \( w \in [0, 1] \) the measurability of the WML estimator can now be obtained by application of Theorem 2.11 of White (1994) or Lemma 2.1 and Theorem 2.2 in Gallant and White (1988).

Proof of Theorem 2. The consistency of the WML estimator can be obtained by appealing to the classical extremum estimation theory found e.g. in Theorem 3.4 of White (1994) or Theorem 3.3 of Gallant and White (1988). In particular, for any weight \( w \in (0, 1] \) and initialization \( \hat{f}_1^* \), the consistency \( \hat{\psi}_T(w, \hat{f}_1^*) \overset{a.s.}{\rightarrow} \psi_0 \) follows from the uniform convergence of the weighted likelihood

\[
\sup_{\psi \in \Psi} |L_T(\psi, w, \hat{f}_1^*) - L_\infty(\psi, w)| \overset{a.s.}{\rightarrow} 0 \forall \hat{f}_1^* \in \mathbb{R}_+ \text{ as } T \rightarrow \infty,
\]

and the identifiable uniqueness of the true parameter \( \psi_0 \in \Psi \) defined e.g. in White (1994).
To establish the uniform convergence of \( \mathcal{L}_T(\psi, w, \tilde{f}_t^*) \) we use the norm sub-additivity inequality

\[
\sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w, \tilde{f}_t^*) - \mathcal{L}_\infty(\psi, w)| \leq \sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w, \tilde{f}_t^*) - \mathcal{L}_T(\psi, w)| + \sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w) - \mathcal{L}_\infty(\psi, w)|
\]

where \( \mathcal{L}_T(\psi, w) \) denotes the likelihood evaluated at the filtered \( \tilde{f}_t(\psi) \) starting in the infinite past. The term

\[
\sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w, \tilde{f}_t^*) - \mathcal{L}_T(\psi, w)|
\]

vanishes by the assumption that \( \|\tilde{f}_t(\psi, w, \tilde{f}_t^*) - \tilde{f}_t(\psi)\|_{\text{a.s.}} \to 0 \), the continuity of the likelihood function and the continuous mapping theorem.

The ergodic theorem for separable Banach spaces of Rao (1962) ensures that

\[
\sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w) - \mathcal{L}_\infty(\psi, w)| \overset{a.s.}{\to} 0
\]

(see also Theorem 2.7 in Straumann and Mikosch (2006)) for the sequence \( \{\mathcal{L}_T(\cdot, w)\} \) of points in \( C(\Psi, \mathbb{R}) \) under:

(i) the SE nature of \( \{\mathcal{L}_T(\cdot, w)\}_{T \in \mathbb{Z}} \) which is ensured by SE nature of \( \{\tilde{f}_t\}_{t \in \mathbb{Z}} \), \( \{x_t\}_{t \in \mathbb{Z}} \) and \( \{y_t\}_{T \in \mathbb{Z}} \), by the continuity of and Proposition 4.3 in Krengel (1985);

(ii) the moment bound \( \mathbb{E}\sup_{\psi \in \Psi} |\ell_t(\psi, w)| < \infty \) ensured by the Gaussian log likelihood under the bounded second moment of \( \tilde{f}_t, x_t \) and \( y_t \).

The identifiable uniqueness of the true parameter \( \psi_0 \in \Psi \), typically defined as

\[
\sup_{\psi : ||\psi - \psi_0|| > \epsilon} \ell_\infty(\psi, w) < \ell_\infty(\psi_0, w) \quad \forall \epsilon > 0
\]

is ensured by the uniqueness of \( \psi_0 \), the compactness of \( \Psi \), and the continuity of \( \mathbb{E}\ell_t(\psi, w) \) on \( \Psi \), which is obtained through the continuity of \( \mathcal{L}_T \) on \( \Psi \) for every \( T \in \mathbb{N} \) and the uniform convergence of the likelihood; see e.g. White (1994). The uniqueness of \( \psi_0 \) as the maximizer \( \mathcal{L}_\infty(\cdot, w) \) for any \( w \in (0, 1) \) is ensured by Theorem 3 which shows that the maximizer \( \mathcal{L}_\infty(\psi_0, w) = 0 \) if and only if \( \mathcal{L}_\infty(\psi_0, 1) = 0 \).

Proof of Theorem 3. The consistency statement follows by the same steps as the proof of Theorem 2 with the exception that the SE nature of \( \{y_t\} \) is assumed rather than derived through Proposition 1.

Let \( z_t \) denote a \( d_z \)-variate random vector with joint density \( p(z_t) \). Furthermore, consider a family of parametric joint densities indexed by the parameter vector \( \psi \), defined as \( Q(\Psi) := \{q(z_t; \psi), \psi \in \Psi\} \). Note that it is possible but not necessary that \( p(z_t) \in Q(\Psi) \). If \( \psi_0^* \) is the pseudo-true parameter that maximizes the limit log likelihood function

\[
\psi_0^* := \arg \max_{\psi \in \Psi} \mathbb{E}_0 \log q(z_t; \psi)
\]

is ensured by the uniqueness of \( \psi_0^* \), the compactness of \( \Psi \), and the continuity of \( \mathbb{E}\ell_t(\psi, w) \) on \( \Psi \), which is obtained through the continuity of \( \mathcal{L}_T \) on \( \Psi \) for every \( T \in \mathbb{N} \) and the uniform convergence of the likelihood; see e.g. White (1994). The uniqueness of \( \psi_0^* \) as the maximizer \( \mathcal{L}_\infty(\psi, w) \) for any \( w \in (0, 1) \) is ensured by Theorem 3 which shows that the maximizer \( \mathcal{L}_\infty(\psi_0, w) = 0 \) if and only if \( \mathcal{L}_\infty(\psi_0, 1) = 0 \).
then it is well known that \( \psi_{0}^{*} \) also minimizes the the Kullback–Leibler divergence \( \text{KL}(p, q(\cdot; \psi)) \) between \( p(z_{t}) \) and \( q(z_{t}; \psi) \) because

\[
\arg\max_{\psi} \mathbb{E}_{0} \log q(z_{t}; \psi) = \arg\min_{\psi} \mathbb{E}_{0} \log p(z_{t}) - \mathbb{E}_{0} \log q(z_{t}; \psi) = \arg\min_{\psi} \text{KL}(p, q(\cdot; \psi)).
\]

Let now the joint density \( q(x_{t}; \psi) \) be factorized into

\[
q_{1}(z_{1,t}|z_{2:d,s,t}; \psi_{0}) \times q_{2}(z_{2:d,s,t}; \psi_{0}) := q_{1}(z_{1,t}|z_{2,t}, \ldots, z_{d,t}; \psi_{0}) \times q_{2}(z_{2,t}, z_{3,t}, \ldots, z_{d,t}; \psi_{0})
\]

and define \( \psi_{0}^{*}(w) \) as the pseudo-true parameter that maximizes the weighted limit log likelihood function

\[
\psi_{0}^{*}(w) := \arg\max_{\psi} \mathbb{E}_{0} \log q_{1}(z_{1,t}|z_{2:d,s,t}; \psi) + w \log \mathbb{E}_{0} q_{2}(z_{2:d,s,t}; \psi).
\]

Then it follows naturally that \( \psi_{0}^{*}(w) \) is the minimizer of the weighted average of KL divergences

\[
\psi_{0}^{*}(w) = \arg\min_{\psi} \text{KL}(q_{1}, p_{1}) + w \text{KL}(q_{2}, p_{2})
\]

because

\[
\arg\max_{\psi} \left[ \mathbb{E}_{0} \log q_{1}(z_{1,t}|z_{2:d,s,t}; \psi) + w \mathbb{E}_{0} q_{2}(z_{2:d,s,t}; \psi) \right]
= \arg\max_{\psi} \left[ \mathbb{E}_{0} \log q_{1}(z_{1,t}|z_{2:d,s,t}; \psi) - \mathbb{E}_{0} \log p_{1}(z_{1,t}|z_{2:d,s,t}) \\
+ w \mathbb{E}_{0} q_{2}(z_{2:d,s,t}; \psi) - w \mathbb{E}_{0} p_{2}(z_{2:d,s,t}) \right]
= \arg\min_{\psi} \left[ \mathbb{E}_{0} \log p_{1}(z_{1,t}|z_{2:d,s,t}) - \mathbb{E}_{0} \log q_{1}(z_{1,t}|z_{2:d,s,t}; \psi) \\
+ w \left( \mathbb{E}_{0} p_{2}(z_{2:d,s,t}) - \mathbb{E}_{0} q_{2}(z_{2:d,s,t}; \psi) \right) \right]
= \arg\min_{\psi} \left[ \text{KL}(q_{1}(\cdot; \psi), p_{1}) + w \text{KL}(q_{2}(\cdot; \psi), p_{2}) \right].
\]

Clearly, if \( w = 1 \), then we obtain the usual ML pseudo-true parameter since

\[
\psi_{0}^{*}(1) = \arg\min_{\psi} \text{KL}(q_{1}(\cdot; \psi), p_{1}) + \text{KL}(q_{2}(\cdot; \psi), p_{2}) = \arg\min_{\psi} \text{KL}(q(\cdot; \psi), p).
\]

For \( w = 0 \) we obtain the ML estimator for the conditional model

\[
\psi_{0}^{*}(0) = \arg\min_{\psi} \text{KL}(q_{1}(\cdot; \psi), p_{1})
\]

Finally, it is also clear that the transformed KL divergence \( \text{TKL}(q, p) := \text{KL}(q_{1}(\cdot; \psi), p_{1}) + w \text{KL}(q_{2}(\cdot; \psi), p_{2}) \) satisfies

\[
\text{KL}(q_{1}(\cdot; \psi), p(\cdot; \psi)) \leq \text{TKL}(q(\cdot; \psi), p) \leq \text{KL}(q(\cdot; \psi), p)
\]
and that $\text{TKL}(q(\cdot; \psi), p)$ is a pre-metric for any $w \in (0, 1]$ as it inherits positivity $\text{TKL}(q(\cdot; \psi), p) \geq 0$ from the positivity of the KL divergence, and satisfies also the identity of indiscernibles since

$$\text{TKL}(q(\cdot; \psi), p) = \text{KL}(q_1(\cdot; \psi), p_1) + w\text{KL}(q_2(\cdot; \psi), p_2) = 0$$

if and only if $q_1(\cdot; \psi) = p_1$ and $q_2(\cdot; \psi) = p_2$, and hence, if and only if $q(\cdot; \psi) = p$. 

**Proof of Theorem 4.** Asymptotic normality of the WML estimator can be obtained by verifying the conditions of Theorem 6.2 of White (1994):

(i) $\hat{\psi}_T(\hat{f}_1^*, w) \xrightarrow{a.s.} \psi_0 \in \text{int}(\Psi)$;
(ii) $\mathcal{L}_T(\cdot, w, \hat{f}_1^*) \in \mathcal{C}^2(\Psi)$ a.s.;
(iii) $\sqrt{T}\mathcal{L}_T'(\psi_0, w, \hat{f}_1^*, \hat{d}\hat{f}_1^*) \xrightarrow{d} N(0, \mathcal{J}(\psi_0, w))$;
(iv) $\sup_{\psi \in \Psi} \|\mathcal{L}_T''(\psi, w, \hat{f}_1^*, \hat{d}\hat{f}_1^*) - \mathcal{L}_T''(\psi, w)\| \xrightarrow{a.s.} 0$;
(v) $\mathcal{L}_T''(\psi, w) = \mathbb{E}\mathcal{L}_T''(\psi, w) = \mathcal{I}(\psi, w)$ is non-singular

Conditions (i) and (ii) follow naturally from Theorem 2, the additional assumption that $\psi_0 \in \text{int}(\Psi)$ and the differentiability of the Gaussian likelihood.

Condition (iii) follows by the asymptotic SE nature of the score $\{\ell_t'(\psi, w, \hat{f}_1^*, \hat{d}\hat{f}_1^*)\}$ which is implied by the SE nature of $\{x_t\}$ and $\{y_t\}$, and the asymptotic SE nature of the filtered sequence $\{\hat{f}_t(\psi, \hat{f}_1^*)\}$ and its derivative $\{\hat{d}\hat{f}_t(\psi, \hat{f}_1^*)\}$. Since the limit score sequence $\{\ell_t'(\psi, w)\}$ is SE we can apply the CLT for SE martingales in Billingsley (1961) to obtain

$$\sqrt{T}\mathcal{L}_T'(\psi_0, w) \xrightarrow{d} N(0, \mathcal{J}(\psi_0, w)) \quad \text{as} \ T \to \infty,$$

where $\mathcal{J}(\psi_0, w) = \mathbb{E}(\ell_t'(\psi_0, w)^T\ell_t'(\psi_0, w)) < \infty$. By Theorem 18.10(iv) in van der Vaart (2000) it thus follows that

$$\sqrt{T}\mathcal{L}_T'(\psi_0, w, \hat{f}_1^*, \hat{d}\hat{f}_1^*) \xrightarrow{d} N(0, \mathcal{J}(\psi_0, w)) \quad \text{as} \ T \to \infty,$$

as long as

$$\|\mathcal{L}_T'(\psi_0, w, \hat{f}_1^*, \hat{d}\hat{f}_1^*) - \mathcal{L}_T'(\psi_0, w)\| \xrightarrow{e.a.s.} 0 \quad \text{as} \ T \to \infty$$

(15) since (15) ensures $\sqrt{T}\|\mathcal{L}_T'(\psi_0, w, \hat{f}_1^*, \hat{d}\hat{f}_1^*) - \mathcal{L}_T'(\psi_0, w)\| \xrightarrow{e.a.s.} 0$ as $T \to \infty$. The e.a.s. convergence in (15) follows from

$$|f_t(\psi_0, w, \hat{f}_1^*) - f_t(\psi_0, w)| \xrightarrow{e.a.s.} 0$$

and

$$\|\hat{d}f_t(\psi_0, \hat{f}_1^*) - \hat{d}f_t(\psi_0)\| \xrightarrow{e.a.s.} 0.$$

Furthermore, since the score of the weighted likelihood is differentiable, we can use the
mean-value theorem to obtain
\[ \| L_T'(\psi_0, w, \tilde{f}_1^*, \tilde{d}f_1^*) - L_T'(\psi_0, w)\| \leq \sum_j |dL_T'| |\tilde{d}f_{j,t}(\psi_0, \tilde{d}f_1^*) - \tilde{d}f_{j,t}(\psi_0)|, \]

where \( \tilde{d}f_{j,t}(\psi_0, \tilde{d}f_1^*) \) denotes the \( j \)-th element of \( \tilde{d}f(\psi_0, \tilde{d}f_1^*) \), and \( dL_T' \) denotes the derivative \( \partial L_T'(\psi_0, w, \tilde{f}_1^*, \tilde{d}f_1^*)/\partial \tilde{d}f_j \) evaluated at some appropriate point between \( \tilde{d}f_{j,t}(\psi_0, \tilde{d}f_1^*) \) and \( \tilde{d}f_{j,t}(\psi_0) \). The bounded moments of the weighted likelihood derivatives and the e.a.s. convergence of the filtered process and its derivatives yield
\[ \| L_T'(\psi_0, w, \tilde{f}_1^*, \tilde{d}f_1^*) - L_T'(\psi_0, w)\| = \sum_j O_p(1) o_{e.a.s.}(1) = o_{e.a.s.}(1). \]

Condition (iv) follows by noting that
\[ \sup_{\psi \in \Psi} \| L_T''(\psi, w, \tilde{f}_1^*, \tilde{d}f_1^*) - L_T''(\psi, w)\| \leq \sup_{\psi \in \Psi} \| L_T''(\psi, w, \tilde{f}_1^*, \tilde{d}f_1^*) - L_T''(\psi, w)\| + \sup_{\psi \in \Psi} \| L_T''(\psi, w) - L_T''(\psi, w)\|. \]

Clearly, \( \sup_{\psi \in \Psi} \| L_T''(\psi, w, \tilde{f}_1^*, \tilde{d}f_1^*) - L_T''(\psi, w)\| \leq 0 \) as \( t \to \infty \) by the continuous mapping theorem and the e.a.s. convergence of the filtered process and its derivatives, and \( \sup_{\psi \in \Psi} \| L_T''(\psi, w) - L_T''(\psi, w)\| \) vanishes by a ULLN under the uniform moment bound on the weighted likelihood \( E \sup_{\psi \in \Psi} \| L_T''(\psi, w)\| \leq \infty \).

Condition (v) is the uniqueness of \( \psi_0 \) as a maximum of \( \ell''_T(\psi, w) \) and ensures the non-singularity of the limit weighted likelihood \( L''_\infty(\psi, w) = E L''_T(\psi, w) = I(\psi, w) \).

**Proof of Theorem 5.** Follows the same steps as the proof of Theorem 4 ith the exception that the required properties of \( \{x_t\} \) and \( \{y_t\} \) are directly assumed rather than derived through Proposition 1.

## B  Low frequency formulations for autoregressions

### B.1  Notation

In this Appendix we discuss the stacking approach similar to the one used by Ghysels (2012) and Foroni, Marcellino, and Schumacher (2015). In the main paper we adopt the stacking method to treat mixed frequency dynamic factor model within a state space setting. But here we provide the details for the autoregressive model case for monthly time series (high frequency) that can be represented by quarterly (low frequency) vectors within a state space formulation. We use the notation \( x_t^m \) for a variable \( x \) that is observed on a monthly \( (m) \) basis with monthly time index \( t \). The observations of the time series \( x_t^m \) can be stacked into a quarterly \( (q) \) observed \( 3 \times 1 \) vector \( x_t^q \) with quarterly time index \( t \).
\[
\begin{pmatrix}
x_{t,1}^q \\
x_{t,2}^q \\
x_{t,3}^q
\end{pmatrix}
= \begin{pmatrix}
x_{3(t-1)+1}^m \\
x_{3(t-1)+2}^m \\
x_{3(t-1)+3}^m
\end{pmatrix},
\]
\hspace{1cm} (19)

where \(x_{t,i}^q\) is the \(i\)-th element of \(x_t^q\) with index \(t\) indicating the number of the quarter of the observation and index \(i\) indicating the number of the month within the quarter; we have \(t = 1, \ldots, T, i = 1, 2, 3\) and \(\tau = 1, \ldots, 3n\), since each quarter consists of \(s = 3\) months.

In a similar way, we can represent the monthly observations into yearly vectors. The monthly observed series \(x_t^m\) can be stacked into a yearly \((y)\) observed \(12 \times 1\) vector \(x_t^y\) with yearly time index \(t\). We then have

\[
\begin{pmatrix}
x_{t,1}^y \\
x_{t,2}^y \\
\vdots \\
x_{t,12}^y
\end{pmatrix}
= \begin{pmatrix}
x_{12(t-1)+1}^m \\
x_{12(t-1)+2}^m \\
\vdots \\
x_{12(t-1)+12}^m
\end{pmatrix},
\]
\hspace{1cm} (20)

where \(x_{t,i}^y\) is the \(i\)-th element of \(x_t^y\) with the first \(t\) indicating the number of the year of the observation and second index \(i\) indicating the number of the month within the year; we have \(t = 1, \ldots, T, i = 1, \ldots, 12\) and \(\tau = 1, \ldots, 12n\), since each year consists of \(s = 12\) months. Throughout this paper, we will use the superscripts \(m, q\) and \(y\) to indicate the frequency of time series; we only use this notation where we deem it is necessary. When a variable or vector has only one index, this index typically refers to the number of the month, quarter or year of the observation. The second index refers to the number of the month within the quarter or year of the observation.

### B.2 Linear State Space Models

The general linear Gaussian state space model can be written in a variety of ways. In this paper we adopt the notation used in Durbin and Koopman (2012), where the model is given as

\[
x_t = Z\alpha_t + \epsilon_t, \quad \epsilon_t \sim NID(0, H),
\]
\[
\alpha_{t+1} = T\alpha_t + R\eta_t, \quad \eta_t \sim NID(0, Q),
\]
\hspace{1cm} (21)

where \(x_t\) is a \(k \times 1\) vector of observations called the *observation vector* and \(\alpha_t\) is an unobserved \(m \times 1\) vector called the *state vector*. The system matrices \(Z, T, R, H\) and \(Q\) are initially assumed to be known and the error terms \(\epsilon_t\) and \(\eta_t\) are assumed to be serially independent and independent of each other at all time points. In practice, some or all of the matrices \(Z, T, R, H\) and \(Q\) will depend on elements of an unknown parameter vector \(\psi\).

In the state space model, the state vector \(\alpha_t\) cannot be observed directly and hence we base the analysis on observations \(x_t\). These equations hold true for any frequency, as
long as the state vector has the same frequency as the observation vector. Therefore, we do not use a superscript to indicate the frequency of the series, although we use these mostly for low frequency models.

The initial state vector \( \alpha_1 \) is generated from \( N(a_1, P_1) \), independently of \( \epsilon_1, \ldots, \epsilon_n \) and \( \eta_1, \ldots, \eta_n \), where \( a_1 \) and \( P_1 \) are assumed known, although \( P_1 \) may depend on the parameter vector \( \psi \).

### B.3 Autoregressive Processes

We now adopt the stacking method in an autoregressive (AR) model setting. For brevity, we illustrate the method using an AR(1). Illustrations with AR(\(p\)) models or higher order \(p\) are provided in Appendix A. Detailed derivations for all models are provided in the Technical Appendix.

Monthly observations from the AR(1) process \( x^m_\tau = \phi x^m_{\tau-1} + \epsilon^m_\tau \) are stacked into the quarterly \(3 \times 1\) vector \( x^q_t \) of (19). The quarterly process of the stacked variable \( x^q_t \) is then given by the vector autoregressive process

\[
x^q_t = Tx^q_{t-1} + R\epsilon^q_t
\]

with

\[
T = \begin{pmatrix} 0 & 0 & \phi \\ 0 & 0 & \phi^2 \\ 0 & 0 & \phi^3 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0 & 0 \\ \phi & 1 & 0 \\ \phi^2 & \phi & 1 \end{pmatrix},
\]

such that the variance matrix of the vector \( x^q_t \), conditional on \( x^q_{t-1} \), is equal to \( \sigma^2 \epsilon RR' \). We notice that all three elements of \( x^q_t \) depend only on the last element of \( x^q_{t-1} \) and on the associating elements of the vector of disturbances \( \epsilon^q_t \). The vector \( \epsilon^q_t \) is the result of stacking the values of \( \epsilon^m_\tau \) in similar fashion as in (19). The autoregressive process (22) is equal to the linear Gaussian state space model (21) with state vector \( \alpha_t = x^q_t \) and with system matrices \( T \) and \( R \) given by (23) and \( Z = I_3, H = 0, Q = \sigma^2 \epsilon \) and \( \eta_t = \epsilon^q_t \).

### B.4 AR(1) process with stacked observations

Consider the AR(1) process of the monthly \((m)\) observed variable \( x^m_\tau \) with monthly time index \( \tau \)

\[
x^m_\tau = \phi x^m_{\tau-1} + \epsilon_\tau, \quad \epsilon_\tau \sim NID \left( 0, \sigma^2 \epsilon \right)
\]

When the monthly observations of \( x^m_\tau \) are stacked into the quarterly \((q)\) \(3 \times 1\) vectors \( x^q_t \) with quarterly time index \( t \), then the equations of the AR(1) process for the stacked observations can be written as

\[
x^q_{t,1} = \phi x^q_{t-1,1} + \epsilon_{t,1}, \quad x^q_{t,2} = \phi x^q_{t-1,2} + \epsilon_{t,2}, \quad x^q_{t,3} = \phi x^q_{t-1,3} + \epsilon_{t,3}
\]
To develop a low frequency recursion for \( x_t^q \), we substitute the first equation for the value of \( x_{t,1}^q \) and the second equation for \( x_{t,2}^q \) gives the following set of equations

\[
\begin{align*}
    x_{t,1}^q &= \phi x_{t-1,3}^q + \varepsilon_{t,1} \\
    x_{t,2}^q &= \phi (\phi x_{t-1,3}^q + \varepsilon_{t,1}) + \varepsilon_{t,2} \\
        &= \phi^2 x_{t-1,3}^q + \phi \varepsilon_{t,1} + \varepsilon_{t,2} \\
    x_{t,3}^q &= \phi (\phi^2 x_{t-1,3}^q + \phi \varepsilon_{t,1} + \varepsilon_{t,2}) + \varepsilon_{t,3} \\
        &= \phi^3 x_{t-1,3}^q + \phi^2 \varepsilon_{t,1} + \phi \varepsilon_{t,2} + \varepsilon_{t,3}
\end{align*}
\]  

(26)

which can be written as the autoregressive process

\[
x_t^q = Tx_{t-1}^q + R\varepsilon_t
\]

(27)

with matrices

\[
T = \begin{pmatrix} 0 & 0 & \phi \\ 0 & 0 & \phi^2 \\ 0 & 0 & \phi^3 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0 & 0 \\ \phi & 1 & 0 \\ \phi^2 & \phi & 1 \end{pmatrix}
\]

(28)

where the variance matrix of the vector \( x_t^q \), conditional on \( x_{t-1}^q \), is equal to \( \sigma^2 \varepsilon RR' \). This autoregressive process is equal to the linear Gaussian state space model

\[
\begin{align*}
    x_t &= Z \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0,H) \\
    \alpha_{t+1} &= T \alpha_t + R \varepsilon_t, \quad \varepsilon_t \sim N(0,Q)
\end{align*}
\]

(29)

with \( Z = I_3 \), with \( \alpha_t = x_t^q \) and \( H = 0 \).

The unconditional variance and covariances (used for initialization of the Kalman Filter) can be obtained from the Yule-Walker equations and are equal to

\[
\begin{align*}
    \gamma_0 &= \frac{\sigma^2 \varepsilon}{(1-\phi^2)} \\
    \gamma_1 &= \phi \gamma_0 \\
    \gamma_2 &= \phi \gamma_1
\end{align*}
\]

(30)

B.5 AR(2) process with stacked observations

For an AR(2) process with monthly observed variable \( x_{t}^m \) the transformations are similar. Consider the model

\[
x_t^m = \phi_1 x_{t-1}^m + \phi_2 x_{t-2}^m + \varepsilon_t, \quad \varepsilon_t \sim NID(0,\sigma^2 \varepsilon)
\]

(31)

The equations of the AR(2) process for the stacked quarterly observations become

\[
\begin{align*}
    x_{t,1}^q &= \phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \varepsilon_{t,1} \\
    x_{t,2}^q &= \phi_1 x_{t-1,1}^q + \phi_2 x_{t-1,2}^q + \varepsilon_{t,2} \\
    x_{t,3}^q &= \phi_1 x_{t-1,2}^q + \phi_2 x_{t-1,1}^q + \varepsilon_{t,3}
\end{align*}
\]

(32)
Substitution of the first equation for the value of \( x_{t,1}^q \) and the second equation for \( x_{t,2}^q \) gives the following set of equations

\[
\begin{align*}
x_{t,1}^q &= \phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \varepsilon_{t,1} \\
x_{t,2}^q &= \phi_1 (\phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,1}^q + \varepsilon_{t,1}) + \phi_2 x_{t-1,2}^q + \varepsilon_{t,2} \\
x_{t,3}^q &= (\phi_1^2 + \phi_2^2) x_{t-1,3}^q + \phi_1 \phi_2 x_{t-1,1}^q + \phi_2 x_{t-1,2}^q + \phi_2 x_{t-1,2}^q + \varepsilon_{t,1} + \varepsilon_{t,3}
\end{align*}
\]

which can be written as the linear Gaussian state space model (29) with matrices

\[
T = \begin{pmatrix} 0 & \phi_2 & \phi_1 \\ 0 & \phi_1 \phi_2 & \phi_1^2 + \phi_2 \\ 0 & \phi_1 \phi_2 + \phi_2^2 & \phi_1^2 + 2 \phi_1 \phi_2 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0 & 0 \\ \phi_1 & 1 & 0 \\ \phi_1^2 + \phi_2 & \phi_1 & 1 \end{pmatrix}
\]

where the variance matrix of the vector \( x_t^q \), conditional on \( x_{t-1}^q \), is equal to \( \sigma_e^2 RR' \). The unconditional variance and covariances are equal to

\[
\begin{align*}
\gamma_0 &= \frac{1 - \phi_2}{1 + \phi_2} \left( \frac{\sigma_e^2}{\phi_1 + \phi_2 - 1} \right) \\
\gamma_1 &= \frac{\phi_1}{(1 - \phi_2) \gamma_0} \\
\gamma_2 &= \phi_1 \gamma_1 + \phi_2 \gamma_0
\end{align*}
\]

### B.6 AR(3) process with stacked observations

For an AR(3) process with monthly observed variable \( x_t^m \) the transformations are again similar. Consider the model

\[
x_t^m = \phi_1 x_{t-1}^m + \phi_2 x_{t-2}^m + \phi_3 x_{t-3}^m + \varepsilon_t, \quad \varepsilon_t \sim NID \left( 0, \sigma_e^2 \right)
\]

The equations of the AR(3) process for the stacked quarterly observations \( x_t^q \) become

\[
\begin{align*}
x_{t,1}^q &= \phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \phi_3 x_{t-1,1}^q + \varepsilon_{t,1} \\
x_{t,2}^q &= \phi_1 x_{t,1}^q + \phi_2 x_{t-1,3}^q + \phi_3 x_{t-1,2}^q + \varepsilon_{t,2} \\
x_{t,3}^q &= \phi_1 x_{t,2}^q + \phi_2 x_{t,1}^q + \phi_3 x_{t-1,3}^q + \varepsilon_{t,3}
\end{align*}
\]

Substitutions similar to those described in the previous subsections can again be applied. We have
where the variance matrix of the process is defined as

\[
\begin{bmatrix}
\sigma_1^2 & \sigma_2^2 & \sigma_3^2 \\
\sigma_2^2 & \sigma_3^2 & \sigma_4^2 \\
\sigma_3^2 & \sigma_4^2 & \sigma_5^2 \\
\end{bmatrix}
\]

which can be written as the linear Gaussian state space model (29) for the quarterly observed \( x_t \) with

\[
T = \begin{pmatrix}
\phi_3 & \phi_2 & \phi_1 \\
\phi_1 \phi_3 & \phi_1 \phi_2 + \phi_3 & \phi_1^2 + \phi_2 \\
\phi_1^2 \phi_3 + \phi_2 \phi_3 & \phi_1 \phi_2 + \phi_3 & \phi_2^3 + 2 \phi_1 \phi_2 + \phi_3
\end{pmatrix}
\]

(39)

For the AR(3) process, \( R \) is the same matrix as for the AR(2) process and the variance matrix of the vector \( x_t \), conditional on \( x_{t-1} \), is again equal to \( \sigma_e^2 RR' \). The three values of \( x_t \) now depend on all three observation of \( x_{t-1} \) and on the 3 x 1 vector of disturbances \( \varepsilon_t \). The unconditional variance and covariances are equal to

\[
\begin{align*}
\gamma_0 &= \frac{\sigma_e^2}{(1-\phi_2-\phi_3-\phi_1^2)(1-\phi_1-\phi_2)(1+\phi_1+\phi_2+\phi_3)} \\
\gamma_1 &= \frac{\sigma_e^2}{(1-\phi_2-\phi_3-\phi_1^2)(1-\phi_1-\phi_2)(1+\phi_1+\phi_2)} \\
\gamma_2 &= \phi_1 \gamma_1 + \phi_2 \gamma_0 + \phi_3 \gamma_1
\end{align*}
\]

(40)

B.7 AR(p) process with stacked observations

For AR(p) processes of order \( p > 3 \) the state vector in (29) has to be extended with more lags of \( x_t \). For example, for the AR(4) process we would have \( \alpha_t = \left( x_{t-1,3}, x_{t-1,1}, x_{t-3,3}, x_{t-3,1} \right)' \) and the transition matrix becomes

\[
T = \begin{pmatrix}
0 & 0 & 0 & 1 \\
\phi_4 & \phi_3 & \phi_2 & \phi_1 \\
\phi_1 \phi_4 & \phi_1 \phi_3 + \phi_4 & \phi_1 \phi_2 + \phi_3 & \phi_1^2 + \phi_2 \\
\phi_1^2 \phi_4 + \phi_2 \phi_4 & \phi_1^2 \phi_3 + \phi_2 \phi_3 + \phi_1 \phi_4 & \phi_1^2 \phi_2 + \phi_1 \phi_3 + \phi_2^2 + \phi_4 & \phi_1^3 + 2 \phi_1 \phi_2 + \phi_3
\end{pmatrix}
\]

(41)

where the variance matrix of the process is defined as \( \sigma_e^2 RR' \) with
\[ R = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \phi_1 & 1 & 0 \\ 0 & \phi_1^2 + \phi_2 & \phi_1 & 1 \end{pmatrix}, \tag{42} \]

The unconditional variances and covariances can be obtained analytically using the Yule-Walker equations, or by numerically solving the Algebraic Riccati Equation.

### B.8 Computing times

<table>
<thead>
<tr>
<th>( p )</th>
<th>Monthly ((n = 12k))</th>
<th>Quarterly ((n = 4k))</th>
<th>Yearly ((n = 1k))</th>
<th>Monthly ((n = 12k))</th>
<th>Quarterly ((n = 4k))</th>
<th>Yearly ((n = 1k))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>13</td>
<td>61</td>
<td>1</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>16</td>
<td>67</td>
<td>2</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>18</td>
<td>76</td>
<td>3</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>41</td>
<td>27</td>
<td>85</td>
<td>4</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>59</td>
<td>40</td>
<td>92</td>
<td>5</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>83</td>
<td>56</td>
<td>100</td>
<td>6</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>106</td>
<td>73</td>
<td>108</td>
<td>7</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>129</td>
<td>90</td>
<td>116</td>
<td>8</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>154</td>
<td>111</td>
<td>124</td>
<td>9</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>191</td>
<td>137</td>
<td>133</td>
<td>10</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>11</td>
<td>226</td>
<td>162</td>
<td>139</td>
<td>11</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>265</td>
<td>190</td>
<td>146</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 8: Computing times. The left panel of this table presents the average computing time (in seconds) that is required to evaluate the loglikelihood function for an AR\((p)\) model of order \(p\) for a monthly time series \(x_{\tau}^{m}\) that is generated by an AR\((p)\) model. Three different approaches are used: treating the data as monthly observations, stacking the data into quarterly \(3 \times 1\) vectors and stacking the data into yearly \(12 \times 1\) vectors. Each value represents the average over 10,000 simulation runs. For each value of \(p\) the fastest of the three approaches is highlighted. The right panel of this table presents the state vector length for each scenario.

### C Interpolation and Aggregation Approaches

Below we discuss the original solution by Mariano and Murasawa (2003) in treating mixed frequency data; they analyze all series at the highest frequency. In this approach, artificial missing values are introduced for the series that is observed at the lower frequency and interpolation techniques are used to describe the dynamics of the unobserved ‘latent’ monthly GDP growth. Furthermore, we explore a second approach where all series are modeled at the lowest frequency by aggregating the high frequency series to quarterly totals. In this approach, no artificial missing values are needed, but information about
the high frequency series is lost and the model does not allow the econometrician to address high frequency dynamics.

C.1 Interpolation Approach

In the mixed frequency interpolation (MFI) approach, all series are treated at the highest frequency, say as monthly time series. All variables are driven by one unobserved monthly common factor $f_\tau$ and by the idiosyncratic factors $u_\tau$ and $v_\tau$. The variable quarterly $y_t$ is observed every third month and has missing values for the first two months of every quarter. Therefore, the data matrix has the following structure

$$
\begin{bmatrix}
\vdots & y_3 & \vdots & y_6 & \cdots & y_{3n} \\
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & \cdots & x_{3n}
\end{bmatrix},
$$

(43)

where $x_\tau$ is the $k \times 1$ vector $(x_\tau^{(1)}, \ldots, x_\tau^{(k)})'$ and $n$ is the number of quarters in the sample period. Since in this approach all series are treated as monthly series, we choose to drop the superscript $m$, in order to avoid cumbersome notation. The contemporaneous and dynamic interactions between $y_\tau$ and the vector of monthly observed variables $x_\tau$ are specified via the model

$$
\begin{bmatrix}
\tilde{y}_\tau^m \\
x_\tau
\end{bmatrix} = \begin{bmatrix}
\beta_y g(f_\tau) \\
\beta_x f_\tau
\end{bmatrix} + \begin{bmatrix}
g(u_\tau) \\
v_\tau
\end{bmatrix} + \begin{bmatrix}
\xi_\tau \\
\varepsilon_\tau
\end{bmatrix}
$$

(44)

where $\tilde{y}_\tau^m$ is the latent monthly variable for $y$ and for which we only have observations available in the last month of each quarter, $\beta_y$ is a scalar coefficient, $\beta_x$ is a $k \times 1$ vector of coefficients, and

$$
g(a_\tau) = \frac{1}{3}a_\tau + \frac{2}{3}a_{\tau-1} + a_{\tau-2} + \frac{2}{3}a_{\tau-3} + \frac{1}{3}a_{\tau-4}
$$

(45)

for $\tau = 1, \ldots, 3n$. The vector $u_\tau$ is the stationary sequence of the idiosyncratic factor for $y_\tau$ and $v_\tau$ is the stationary sequence of the vector of idiosyncratic factors for $x_\tau$ and consists of one value per month for each monthly observed variable $x_\tau^{(i)}$.

The factors $f_\tau$, $u_\tau$ and $v_\tau$ are modeled as AR processes

$$
f_\tau \sim AR(p_f), \quad u_\tau \sim AR(p_u), \quad v_\tau \sim AR(p_v)
$$

(46)

where $AR(p)$ refers to the autoregressive process of order $p$. There are no interactions between the series of $f_\tau$, $u_\tau$ and $v_\tau$ nor between the series of $v_\tau^{(i)}$ and $v_\tau^{(j)}$ for any $i \neq j$.

We have established the model by Mariano and Murasawa (2003) and implicitly their model-based solution to the mixed frequency problem. They advocate to use the Kalman filter for likelihood evaluation and general analysis. Specifically, they take advantage of the fact that the Kalman filter can treat missing observations without a problem.


C.2 Aggregation Approach

An alternative approach, where the introduction of artificial missing values is not required, is the aggregation of the monthly series \( x_{m_\tau} \) into quarterly totals \( \bar{x}_q^t \) and the treatment of all series as quarterly series. This model then only describes quarterly dynamics. The unobserved common factor must also become quarterly. To avoid cumbersome notation we drop the superscript \( q \) and we write the model in the form

\[
\begin{pmatrix}
    y_t \\
    \bar{x}_t
\end{pmatrix} = \begin{bmatrix}
    \beta_y & 1 & 0 \\
    \beta_x & 0 & I_k
\end{bmatrix} \begin{pmatrix}
    f_t \\
    u_t \\
    \bar{v}_t
\end{pmatrix} + \begin{pmatrix}
    \xi_t \\
    \bar{\epsilon}_t
\end{pmatrix},
\]

for \( t = 1, \ldots, T \), where \( \bar{x}_t \) and \( \beta_x \) are both \( k \times 1 \) vectors. The common factor \( f_t \) and the idiosyncratic factors \( u_t \) and \( \bar{v}_t \) can still be modeled by AR processes as in (46). However, we must take care when interpreting the values of the parameters of these processes, as they now describe the dynamics from quarter to quarter. We will hereafter refer to this approach as the mixed frequency aggregation (MFA) approach.