Journal of Econometrics 193 (2016) 405–417

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Weighted maximum likelihood for dynamic factor analysis and forecasting with mixed frequency data

F. Blasques [a,b], S.J. Koopman [a,b,c,*], M. Mallee [a], Z. Zhang [a,b]

[a] Vrije Universiteit Amsterdam, Netherlands
[b] Tinbergen Institute Amsterdam, Netherlands
[c] CREATES, Aarhus University, Denmark

ARTICLE INFO

Article history:
Available online 6 May 2016

JEL classification:
C13
C32
C53
E17

Keywords:
Asymptotic theory
Kalman filter
Nowcasting
State space

ABSTRACT

For the purpose of forecasting key macroeconomic or financial variables from a panel of time series variables, we adopt the dynamic factor model and propose a weighted likelihood-based method for parameter estimation. 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 may contribute to the forecasting of 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 verify the new estimation method in a Monte Carlo study and investigate the role of different weights in different settings. In the context of 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. We adopt a low frequency representation of the mixed frequency dynamic factor model and discuss the computational efficiencies of this approach. In our empirical study for the U.S. economy, we present improvements in nowcasting and forecasting accuracy when the weighted likelihood-based estimation procedure is adopted.

1. Introduction

The forecasting of macroeconomic and financial time series variables is of key importance for economic policy makers. Reliable forecasts are especially in high demand when the economic environment is uncertain as we have witnessed in the years during and after the financial crisis. 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 variables of interest and related variables. The dynamic factor model has become a popular tool for the 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 dynamic factor model can be viewed as a high-dimensional linear state space model. The estimation of the parameters in a dynamic factor model is a challenging task given the large number of parameters, mostly due to factor loading coefficients. A likelihood-based approach in which the Gaussian likelihood function is evaluated via the Kalman filter and is numerically maximized with respect to the parameter vector has been originally proposed by Engle and Watson (1981) for a model with one dynamic factor. Watson and Engle (1983) base their estimation procedure on an expectation–maximization (EM) algorithm; see also Quah and Sargent (1993). More recently, feasible two-step approximate likelihood-based procedures are developed by Doz et al. (2011) and Bańbura and Modugno (2014). In Bräuning and Koopman (2014) and Jungbacker and Koopman (2015), specific data transformations are considered to facilitate the parameter estimation for high-dimensional dynamic factor models. In this study, we restrict ourselves to likelihood-based estimation procedures.

* Correspondence to: Department of Econometrics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081HV Amsterdam, The Netherlands.
E-mail addresses: f.blasques@vu.nl (F. Blasques), s.j.koopman@vu.nl (S.J. Koopman), maxmallee@hotmail.com (M. Mallee), z2.zhang@vu.nl (Z. Zhang).

http://dx.doi.org/10.1016/j.jeconom.2016.04.014
0304-4076/

Weighted likelihood-based estimation

To address the notion that a single variable or a small selection of variables in a dynamic factor model is of key importance while all other variables can be regarded as instruments, we present a weighted likelihood-based estimation procedure for the purpose of providing a more accurate forecasting performance than obtained from a standard maximum likelihood procedure. Our proposed weighted maximum likelihood estimator gives simply more weight to the likelihood contribution from the variable of interest. As an example, for the nowcasting and forecasting of quarterly growth in gross domestic product, referred to as GDP growth, more weight can be given to the likelihood contribution from GDP growth in comparison to the contribution from the related variables that are included in the dynamic factor model.

The variable-specific weights introduced by our weighted ML estimator differ from other weighted ML estimators proposed in the literature. In most other cases, observation-specific weights in the likelihood function are considered. 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 et al. (1997, 1998) down-weights observations that are inconsistent with the postulated model. Similarly, Hu and Zidek (unpublished) 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 et al. (2004) for given weights, and Wang and Zidek (2005) provide a method for estimating the weights based on cross-validation. In contrast we propose a weighted ML estimator that gives higher weight to a subset of a random vector, that is to an entire random scalar sequence within the multivariate stochastic process.

We discuss the asymptotic properties of our weighted maximum likelihood estimator and we show that the estimator is consistent and asymptotically normal. We also verify our new approach in a Monte Carlo study to investigate the effect of different choices for the weights in different scenarios. In an empirical study concerning the nowcasting and forecasting of U.S. GDP growth, we adopt the weighted likelihood function for the estimation of parameters in a mixed frequency dynamic factor model.

Mixed Frequency

In empirical studies, the dynamic factor model requires further modifications to handle mixed frequency data; Mariano and Murasawa (2003) have been the first to illustrate how a small-scale dynamic factor model for the U.S. economy can be adapted for mixed frequency data. Their model is formulated in state space form with a monthly time index. The monthly and quarterly variables are dependent on a common monthly dynamic factor and on idiosyncratic dynamic components. For the quarterly variable of interest, the Kalman filter can treat the missing observations that occur during the first two months in each quarter. More generally, any multivariate time series model can be formulated in terms of a high frequency time index and the periodically missing observations due to low frequency variables can be accounted for by the Kalman filter. Mittnik and Zadrozny (2005) report promising results based on this approach for the forecasting of German growth in GDP.

We consider an alternative approach based on ideas developed for periodic systems in the control engineering literature; see Bittanti and Colaneri (2000, 2009) The main idea is to formulate the model with a low frequency time index and collect the observations for a high frequency variable in a vector. In the case of a quarterly time index and a monthly variable, the three consecutive monthly observations associated with a specific quarter are then stacked into a quarterly vector. Both monthly and quarterly dynamic processes can be formulated in a state space model with a quarterly time index. We discuss this solution for the mixed frequency dynamic factor model. The advantage of this approach is that it does not require the handling of missing observations and it can lead to computational efficiencies. A similar solution is considered by Marcellino et al. (2014) who propose a Bayesian regression model with stochastic volatility for producing current-quarter forecasts of GDP growth using many monthly economic variables. Such ideas are also explored for vector autoregressive systems by Chen et al. (2012), Ghysels (2012), Foroni et al. (2015) and Ghysels et al. (forthcoming).

Empirical study

An important application of dynamic factor models is their use in the forecasting of quarterly GDP growth. A 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 (2002b) and Giannone et al. (2008) for the U.S., Marcellino et al. (2003) and Rünstler et al. (2009) for the euro area, and Schumacher and Breitung (2008) for Germany. In many of these and related 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 a highly relevant issue in forecasting, nowcasting and backcasting GDP growth; see also the discussions in Bańbura et al. (2013).

In our empirical study for the U.S. economy, we consider three small- to medium-sized mixed frequency dynamic factor models with the purpose of forecasting quarterly U.S. GDP growth. The first model is a five-dimensional model similar to Mariano and Murasawa (2003), the second model is a fourteen-dimensional model similar to Bańbura et al. (2013) and the third model is a six-dimensional model similar to Aruoba et al. (2009). The first two models have only monthly related variables while the last model also includes a weekly related variable. For almost all cases, we present improvements in nowcasting and forecasting accuracy when parameters are estimated by the weighted maximum likelihood method.

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 low frequency formulations lead to computational gains. In Section 4 we present and explore the results of our empirical study concerning U.S. GDP growth. We compare the nowcasting and forecasting accuracies of our new approach for the three different dynamic factor models. We also establish the empirical relevance of the weighted estimation method of Section 2. Section 5 summarizes and concludes.

2. Weighted maximum likelihood: method and properties

We represent our high-dimensional panel of time series as the column vector z_t for which we have observations from $t = 1, \dots, T$ where T is the overall time series length. 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

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

for $t = 1, \ldots, T$, where $\Lambda = [\Lambda_y', \ \Lambda_x']'$ is the factor loading matrix with dimension $N_z \times r$, Λ_i is the $N_i \times r$ factor loading sub-matrix Λ_i, for $i = y, x$ with A' being the transpose of matrix A, f_t is the $r \times 1$ vector with latent dynamic factors, $\varepsilon_t = (\varepsilon_{y,t}', \varepsilon_{x,t}')'$ is the observation disturbance vector, with $\varepsilon_{i,t}$ as the normally distributed $N_i \times 1$ disturbance vector, for $i = x, y$, Φ is the autoregressive coefficient matrix, and η_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 Φ 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 $\mathrm{E}(f_t) = 0$ and $\mathrm{Var}(f_t) = I_r$ 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 ε_t and η_t are assumed to be mutually and serially uncorrelated for all time periods. In particular, we have

$$\varepsilon_t \sim N(0, \Sigma_\varepsilon), \qquad \eta_t \sim N(0, \Sigma_\eta), \qquad \mathrm{Cov}(\varepsilon_t, \eta_s) = 0, \tag{2}$$

for $t, s = 1, \ldots, T$. To enforce the normalization of the factors f_t, the variance matrix Σ_η is restricted to be $\Sigma_\eta = I_r - \Phi\Phi'$. The remaining coefficient matrices Λ, Σ_ε and Φ are functions of the unknown parameter vector that we denote by ψ. 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 the mixed frequency dynamic factor model that we adopt in Section 3.

Different methods have been proposed for the estimation of the unknown parameter vector ψ; see the discussion in the introductory section. We restrict ourselves to those methods that are likelihood-based and aim at maximizing the loglikelihood function. The loglikelihood function relies on the joint log density $\log p(z; \psi) = \log p(y, x; \psi)$ 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 ψ. It is a standard exercise of numerical optimization in which the Kalman filter evaluates the loglikelihood function whenever a different ψ 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 (1)–(2) provides a convenient framework for obtaining simple descriptions of potentially complex interactions between the economic variables. In particular, the common factors summarize partly the commonalities 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 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 is not necessarily sufficient to abandon the ML estimator, but taken together, they are. 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 good reasons to employ the ML estimator. Under standard 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 pseudo-true parameter optimizes forecasting accuracy. However, we will argue that when we take the above points (i) and (ii) together, the ML estimator is no longer the best possible estimator available when interest lies in the forecasting of only a subset of the observed variables. These characteristic features call for a novel estimation procedure with the aim to improve the forecasting accuracy of the variable of interest in the context of a misspecified model. We provide both theoretical and simulation-based evidence that a weighted ML estimator outperforms the classic ML estimator in forecasting the variable of interest. In Section 4, we show that the ability to outperform the ML estimator is also visible in an empirically relevant application to economic forecasting.

2.1. Weighted maximum likelihood estimator

Consider the dynamic factor model (1) and (2) where both y_t and x_t can be treated as vectors. The loglikelihood function for the model is given by

$$\mathcal{L}_T(\psi, f_1^*) := \log p(y, x; \psi) = \log p(y|x; \psi) + \log p(x; \psi), \tag{3}$$

where the initial value of the factor $f_1 = f_1^*$ and the parameter vector ψ 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; see Koopman and Durbin (2000) for a computationally efficient implementation in the context of a large-dimensional observation vector $z_t = (y_t', x_t')'$.

The maximum likelihood estimation of parameter vector ψ is based on applying a numerical quasi-Newton optimization method for the maximization of $\mathcal{L}_T(\psi, f_1^*)$, with respect to ψ. The numerical maximization is an iterative process. After its convergence, the maximum likelihood estimate of ψ 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 our empirical study in Section 4, 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 most 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. We do not want to disregard the likelihood contribution of x_t completely since we still require an appropriate model representation of x_t given the dependence of y_t on x_t through the common dynamic factor f_t. 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

$$\mathcal{L}_T(\psi, w, f_1^*) = W \cdot \log p(y|x;\psi) + \log p(x;\psi), \tag{4}$$

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 ψ that maximizes (4) is referred to as the weighted maximum likelihood (WML) estimator.

The novel WML estimator differs from other weighted ML estimators proposed in the literature. Our 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 et al. (1997, 1998) is 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 (unpublished) assigns different weights to different observations in the likelihood function.

The motivation for the development of our WML estimator is also different. The 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.

2.2. 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 pre-metric that gives more weight to fitting the conditional density of y_t when $0 < w < 1$. We use the term pre-metric to identify a map that satisfies all the axioms of metrics except for the *axiom of symmetry*, for example $d(x, y) = d(y, x)\ \forall\ (x, y)$. 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 the propositions and theorems presented in this section below are presented in the Online Appendix.

Proposition 1 states standard 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. Below, we let $\|\cdot\|$ denote the Euclidean spectral norm, i.e. the p-norm that sets $p = 2$. The results extend naturally to other norms.

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

(i) $\|\Phi\| < 1$ *in (1) and* $0 < \|\Sigma_\eta\| < \infty$ *in (2);*
(ii) $\|\Lambda_x\| < \infty$ *in (1) and* $0 < \|\Sigma_\varepsilon\| < \infty$ *in (2);*
(iii) $\|\Lambda_y\| < \infty$ *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 $\hat{\psi}_T(w, \tilde{f}_1^*) : \Omega \to \Psi$ satisfying*

$$\hat{\psi}_T(w, \tilde{f}_1^*) \in \arg\max_{\psi\in\Psi} \mathcal{L}_T(\psi, w, \tilde{f}_1^*),$$

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. Since the true time-varying parameter $\{f_t\}_{t\in\mathbb{Z}}$ is unobserved, the estimation of the true parameter vector ψ_0 depends crucially on the use of a filter for this sequence. Below, we use tildes to distinguish the filter from the true parameter and hence let $\tilde{f}_t(\psi, \tilde{f}_1^*(\psi))$ denote the filter at time t, parameterized by ψ and initialized at point $\tilde{f}_1^*(\psi)$ at time $t = 1$. The asterisk highlights that $\tilde{f}_1^*(\psi)$ is simply an initialization for the filter. Note that the filter depends on the vector of static parameters $\psi \in \Psi$. Naturally, it is the static parameter ψ that defines the properties of the filtering sequence $\{\tilde{f}_t(\psi, \tilde{f}_1^*(\psi))\}_{t\in\mathbb{N}}$.

From an estimation perspective, we are interested in the filtered sequence as a function of the vector of parameters ψ. As such, we will adopt the stochastic recurrence approach of Straumann and Mikosch (2006) and concentrate on the random functions $\tilde{f}_t(\tilde{f}_1^*) = \tilde{f}_t(\cdot, \tilde{f}_1^*(\cdot))$ that take values in the separable Banach space $\mathbb{C}(\Psi, \|\cdot\|_\Psi)$ of continuous functions defined on Ψ, equipped with supremum norm $\|\cdot\|_\Psi$.

The consistency of the WMLE established in Theorem 2 is obtained under the assumption that the common factor model is correctly specified. Furthermore, Theorem 2 holds for any filter that identifies the parameter vector $\psi_0 \in \Psi$ and is asymptotically SE with bounded moments of second order. The identification of $\psi_0 \in \Psi$ is naturally ensured when the filter is invertible since then the limit filtering sequence $\{\tilde{f}_t\}_{t\in\mathbb{Z}}$ will then satisfy $\tilde{f}_t(\psi_0) = f_t\ \forall\ t \in \mathbb{Z}$. Furthermore, the exponential almost sure (e.a.s.) convergence of the filter $\{\tilde{f}_t(\tilde{f}_1^*)\}_{t\in\mathbb{N}}$ to a limit SE sequence $\{\tilde{f}_t\}_{t\in\mathbb{Z}}$ that does not depend on the initialization $\tilde{f}_1^*$ is ensured as long as standard contraction conditions hold. Both the identification condition and the e.a.s. convergence of the filter to an SE process with bounded second moment are standard and easy to establish in this linear Gaussian setting. For this reason, we do not repeat them here; see e.g. Mehra (1970) and Bougerol and Picard (1992) for such results on the classical Kalman filter, Bougerol (1993) and Straumann and Mikosch (2006) for extensions to other filters, and Blasques et al. (2014) for identification, convergence results and bounded moments on a wide range of observation-driven filters. Theorem 2 thus assumes that ψ_0 maximizes the likelihood and assumes the convergence of the filtered sequence $\{\tilde{f}_t(\tilde{f}_1^*)\}_{t\in\mathbb{N}}$ initialized at

$\tilde{f}_1^* \in \mathbb{C}(\Psi)$ to a unique limit SE sequence $\{\tilde{f}_t\}_{t\in\mathbb{Z}}$ with bounded second moment.

The uniform convergence of the filtering sequence to an SE limit lets us to establish the uniform convergence of the weighted loglikelihood function by application of a uniform law of large numbers. We denote the limit weighted loglikelihood function by $\mathcal{L}_\infty(\psi, w)$. We only require the identification of ψ_0 in the usual ML setting ($w = 1$); that is identification w.r.t. the unweighted loglikelihood function $\mathcal{L}_T(\psi, 1)$. In the proof, we show that identification of ψ_0 in $\mathcal{L}_T(\psi, 1)$ implies identification of ψ_0 in $\mathcal{L}_T(\psi, w)$, 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 Proposition 1 and Theorem 1 hold. Suppose furthermore that*

$$\mathcal{L}_\infty(\psi_0, 1) > \mathcal{L}_\infty(\psi, 1) \quad \forall\, \psi \neq \psi_0$$

and there exists a unique SE sequence such that

$$\|\tilde{f}_t(\tilde{f}_1^*) - \tilde{f}_t\|_\Psi \xrightarrow{e.a.s.} 0 \quad \forall\, \tilde{f}_1^* \text{ as } t \to \infty \text{ with } \mathbb{E}|\tilde{f}_t|^2 < \infty.$$

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 weight $w \in (0, 1]$.

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 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

$$\|\tilde{f}_t(\tilde{f}_1^*) - \tilde{f}_t\|_\Psi \xrightarrow{e.a.s.} 0 \quad \text{for every initialization } \tilde{f}_1^* \text{ as } t \to \infty \text{ with } \mathbb{E}|\tilde{f}_t|^2 < \infty.$$

Then

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

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

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

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

$$\text{TKL}_1\big(q(\cdot;\psi), p\big) = \text{KL}\big(q(\cdot;\psi), p\big),$$
$$\text{TKL}_0\big(q(\cdot;\psi), p\big) = \text{KL}\big(q_1(\cdot;\psi), p_1\big),$$
$$\text{KL}\big(q_1(\cdot;\psi), p_1\big) \le \text{TKL}_w\big(q(\cdot;\psi), p\big) \le \text{KL}\big(q(\cdot;\psi), p\big),$$
$$\text{and} \quad \text{TKL}_w\big(q(\cdot;\psi), p\big) = 0 \quad \text{if and only if} \quad \text{KL}\big(q_1(\cdot;\psi), p_1\big) = 0.$$

Theorem 4 establishes the asymptotic normality of the WML estimator under the assumption that the dynamic factor model is correctly specified. Below we let $\mathcal{J}(\psi_0, w) := (\partial\mathbb{E}\ell_t(\psi_0, w)/\partial\psi) \times (\partial\ell_t(\psi_0, w)/\partial\psi)'$ denote the expected outer product of gradients and $\mathcal{I}(\psi_0, w) := \partial^2\mathbb{E}\ell_t(\psi_0, w)/\partial\psi\partial\psi'$ 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, $\{\tilde{df}_t(\tilde{df}_1^*)\}$ and $\{\tilde{ddf}_t(\tilde{ddf}_1^*)\}$ denote the first and second derivatives of the filter w.r.t. the parameter vector ψ, initialized at $\tilde{df}_1^*$ and $\tilde{ddf}_1^*$, respectively. Their SE limits are denoted $\{\tilde{df}_t\}$ and $\{\tilde{ddf}_t\}$. 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 ψ_0 be a point in the interior of Ψ. Suppose furthermore that there exists a unique SE sequence $\{\tilde{df}_t\}$ such that*

$$\|\tilde{df}_t(\tilde{df}_1^*) - \tilde{df}_t\|_\Psi \xrightarrow{e.a.s.} 0 \quad \forall\, \tilde{df}_1^* \text{ as } t \to \infty \text{ with } \mathbb{E}|\tilde{df}_t|^4 < \infty$$

and a unique SE sequence $\{\tilde{ddf}_t\}$ such that

$$\|\tilde{ddf}_t(\tilde{ddf}_1^*) - \tilde{ddf}_t\|_\Psi \xrightarrow{e.a.s.} 0 \quad \forall\, \tilde{ddf}_1^* \text{ as } t \to \infty \text{ with } \mathbb{E}|\tilde{ddf}_t|^2 < \infty.$$

Then, for every $\tilde{f}_1^*$ and every $w \in (0, 1]$, the ML estimator $\hat{\psi}_T(\tilde{f}_1^*)$ satisfies

$$\sqrt{T}\big(\hat{\psi}_T(\tilde{f}_1^*, w) - \psi_0\big) \xrightarrow{d} N\Big(0,\ \mathcal{I}^{-1}(\psi_0, w)\mathcal{J}(\psi_0, w)\mathcal{I}^{-1}(\psi_0, w)\Big) \quad \text{as } T \to \infty.$$

Naturally, we can extend the asymptotic normality results to the misspecified 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 Ψ. 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 $\{\tilde{df}_t\}$ such that*

$$\|\tilde{df}_t(\tilde{df}_1^*) - \tilde{df}_t\|_\Psi \xrightarrow{e.a.s.} 0 \quad \forall\, \tilde{df}_1^* \text{ as } t \to \infty \text{ with } \mathbb{E}|\tilde{df}_t|^4 < \infty$$

and a unique SE sequence $\{\tilde{ddf}_t\}$ such that

$$\|\tilde{ddf}_t(\tilde{ddf}_1^*) - \tilde{ddf}_t\|_\Psi \xrightarrow{e.a.s.} 0 \quad \forall\, \tilde{ddf}_1^* \text{ as } t \to \infty \text{ with } \mathbb{E}|\tilde{ddf}_t|^2 < \infty.$$

Then, for every $\tilde{f}_1^*$ and every $w \in (0, 1]$, the ML estimator $\hat{\psi}_T(w, \tilde{f}_1^*)$ satisfies

$$\sqrt{T}\big(\hat{\psi}_T(\tilde{f}_1^*) - \psi_0^*(w)\big) \xrightarrow{d} N\Big(0,\ \mathcal{I}^{-1}(\psi_0^*(w), w)\mathcal{J}(\psi_0^*(w), w)\mathcal{I}^{-1}(\psi_0^*(w), w)\Big)$$

as $T \to \infty$.

2.3. 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 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 into 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 ψ 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 in Box I.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 ($\text{MSE}_1(w)$)

$$\hat{w}_H = \arg\min_{w\in[0,1]} \frac{1}{H}\sum_{i=1}^{H} e_{+1}^i(w)^2 = \arg\min_{w\in[0,1]} \text{MSE}_1(w).$$

Naturally, the criterion can be easily redesigned for w to minimize the h-step ahead forecast error (MSE_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 MSE 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 MSE 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*

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

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

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

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).

2.4. Small sample properties of WML: a Monte Carlo study

In our Monte Carlo study, we investigate the finite sample effects of different choices for the value of W on the in-sample fit for two different data generation processes (DGPs). The first DGP for $z_t = (y_t', x_t')'$ is the dynamic factor model (1)–(2) and the second DGP is a stationary vector autoregressive (VAR) model. The panel dimension for y_t is $N_y = 1$ (we have a univariate time series y_t) and we consider three different panel dimensions for x_t, specifically $N_x = 2$, $N_x = 5$ and $N_x = 10$. The time series length in all simulations is set to $T = 120$. The first 80 observations are used for parameter estimation while the remaining 40 observations are used for one-step ahead forecast evaluation. The mean squared error (MSE) of the 40 forecast errors of the univariate time series y_t is used for the selection of the optimal value for W in the WML method. In our Monte Carlo study we only consider these 40 "in-sample" forecasts.

The first DGP is the DFM (1)–(2) with a single dynamic factor f_t, that is $r = 1$. The parameter matrices of the DFM are chosen as follows: the factor loading matrix is given by $\Lambda = (1, 1, 1/2, 1/3, \ldots, 1/N_x)'$, the diagonal elements of Σ_ε are set to 0.5 and the persistence coefficient for the single factor f_t is given by where $\Sigma_\varepsilon = 0.8$. The second DGP for z_t is the stationary vector autoregressive process of order 1, VAR(1), and is given by $z_t = \Psi z_{t-1} + \xi_t$ with normally independently distributed $N_z \times 1$ disturbance vector ξ_t, that has mean zero and variance matrix Σ_ξ, and with $N_z \times N_z$ autoregressive coefficient matrix Ψ. The stationary conditions with respect to Ψ apply. The parameter matrices of the VAR(1) are chosen as follows: the matrix Ψ is an upper-triangular matrix with its diagonal elements equal to 0.8, its upper-triangular elements are randomly and uniformly chosen within the range $(-0.5, 0.5)$, and the variance matrix Σ_ξ is set equal to $0.5 \cdot I_{N_z}$. These choices for the parameters imply a stationary VAR(1) model.

In our simulation experiment, we consider two different settings of correct specification and misspecification. In the first case, we adopt the DFM as the DGP and we consider the same model for estimation and forecasting. We generate three databases of 500 vector time series z_t with $N_y = 1$, and each with $N_x = 2$,

$$\begin{array}{cccccccccccc}
y_1 & y_2 & y_3 & \cdots & \cdots & y_{T'} & \boxed{\hat{y}_{T'+1}} & & & & \\
 & y_2 & y_3 & \cdots & \cdots & y_{T'} & y_{T'+1} & \boxed{\hat{y}_{T'+2}} & & & \\
 & & & & \cdots & & & \cdots & & & \\
 & & & y_H & \cdots & y_{T'} & y_{T'+1} & y_{T'+2} & \cdots & y_{T'+H} & \boxed{\hat{y}_{T'+H+1}}.
\end{array}$$

Box I.

$N_x = 5$ and $N_x = 10$. For each simulated vector time series z_t, we estimate the parameters, using the first 80 observations, via the WML method for a range of values for W, including $1, \ldots, 5$, 10, 50 and 100. Notice that for $W = 1$, the WML method reduces to the ML method. Based on the parameter estimates and on the next 40 observations, we compute the mean squared error (MSE) of the one-step ahead forecast errors for the target variable y_t. The forecast errors are obtained from the Kalman filter. Under correct specification, we expect that increasing the value of W will not improve the forecasting accuracy for the variable of interest y_t. Theorem 2 implies 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 second case, the VAR(1) model is the DGP as described above while the DFM is considered for estimation and forecasting. Similarly as in the first case, we generate three databases of 500 vector time series from the VAR(1) process, each with $N_z = 3$, $N_z = 6$ and $N_z = 11$. Subsequently, we estimate the DFM parameters with the values of W as given above. In the misspecification case, we also consider the W values 250, 500 and 1000, since we expect that an increasing weight W will be beneficial for the forecasting accuracy of y_t. Theorem 3 implies 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.

In Table 1, for each entry, we present the average of the MSE for the 500 generated time series of the target variable y_t, scaled by the average MSE obtained by the ML method, for the corresponding case, that is $W = 1$. From the misspecification case (right-hand side panel), we learn that increasing W leads to a better in-sample forecasting accuracy for y_t, for all three dimensions N_x. It is not necessary to choose a very large value of W. The improvements in the average MSE for larger W appear to converge to some upper limit, for all three dimensions N_x. Furthermore, we observe that more gains are made when more variables are included in the model so that the misspecification is more pronounced as the number of factors remains one. The correct specification case (left-hand side panel of Table 1) reveals that 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 $N_x = 5$, while the improvement is only about 0.11% compared to the benchmark of $W = 1$. Overall, a larger value for W does not lead to a better forecasting accuracy for y_t when the model is correctly specified.

In panel I of Table 2, we present the frequencies of optimal W values in 500 Monte Carlo simulations. The results confirm the findings from Table 1. Under correct specification, the weights W that are close to unity provide 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 panel II of Table 2 we also report the sample rejection rates of the Diebold–Mariano test at 90% confidence level over the 500 simulations for the forecasts obtained from WML parameter estimates against the forecasts obtained from ML parameter estimates. Under correct model specification, the rejection rate can be viewed as the size of the DM test in our setting. While under misspecification, the rejection rate can be viewed as the power of the DM test. For $N_x = 2$, the rejection rate of the DM test is 4.4% and the test is clearly undersized. However, for $N_x = 5$ and $N_x = 10$, the rejection rates are 8.4% and 11.8%, respectively. The results roughly suggest that we obtain the correct size when N_x gets larger. On the other hand, the rejection rate of the DM test is 91.4% when $N_x = 2$ in the misspecification case. The rejection rate increases when we include more variables in x_t. These results indicate that the power of the test is large when N_x gets larger. Hence we support the conclusion from Stock and Watson (2002a) that large-scale dynamic factor models are preferred 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 1
Monte Carlo results I. In-sample one-step ahead forecasting of target variable y_t using the dynamic factor model with parameter estimates obtained from the WML method for different values of W. We report the mean squared error (MSE), averaged over 500 Monte Carlo simulations. The time series sample size is $T = 120$, a single target variable is considered, $N_y = 1$, and the panel dimensions for x_t are $N_x = 2, 5, 10$. In the case of 'correct specification', the DFM model (1)–(2), with $r = 1$, $\Lambda_y = 1$, $\Lambda_x = (1, 1/2, \ldots, 1/N_x)'$, $\Sigma_\varepsilon = 0.5 I_{N_z}$ and $\Phi = 0.8$, is used for data generation of y_t and x_t. In the case of 'misspecification', y_t and x_t are generated by the stationary vector autoregressive model $z_t = \Psi z_{t-1} + \xi_t$, with diagonal elements of Ψ equal to 0.8, its lower-diagonal elements are set to zero, its upper-diagonal elements are uniformly sampled within the range $(-0.5, 0.5)$, and the variance matrix of the zero-mean disturbance vector ξ_t is a diagonal matrix with its diagonal elements equal to 0.5. In both cases, the forecasts are obtained from the DFM as described above with estimated Λ, Σ_ε (diagonal) and Φ by the WML method. The smallest average MSE in each column is highlighted.

W	Correct specification			Misspecification		
	$N_x = 2$	$N_x = 5$	$N_x = 10$	$N_x = 2$	$N_x = 5$	$N_x = 10$
	I: In-sample forecast mean squared errors					
1	**1.0000**	1.0000	1.0000	1.0000	1.0000	1.0000
2	1.0119	0.9993	0.9964	0.9527	0.9439	0.9310
3	1.0124	**0.9989**	**0.9960**	0.9281	0.9081	0.8870
4	1.0134	1.0001	0.9961	0.9105	0.8806	0.8551
5	1.0141	1.0006	0.9974	0.8981	0.8538	0.8294
6	1.0150	1.0002	0.9972	0.8872	0.8329	0.8097
10	1.0190	1.0023	1.0001	0.8578	0.7794	0.7572
15	1.0224	1.0040	1.0024	0.8384	0.7388	0.7012
25	1.0267	1.0064	1.0043	0.8177	0.6914	0.6547
50	1.0319	1.0092	1.0062	0.7955	0.6330	0.5862
100	1.0367	1.0110	1.0084	0.7751	0.5930	0.5056
250				0.7648	0.5530	0.4551
500				0.7561	0.5359	0.4168
1000				**0.7511**	**0.5209**	**0.3880**

3. Mixed frequency dynamic factor model

In empirical forecasting studies based on dynamic factor analyses, we typically analyze variables that are observed at different frequencies. More specifically, and most relevant for economic forecasting, we consider settings for which the variable of interest y_t is observed at a low frequency, quarter by quarter, while the related variables in x_t are observed at high frequencies such as monthly, weekly and daily. To accommodate mixed frequency time series panels, we need to modify the dynamic factor model (1)–(2) appropriately. We will discuss a number of solutions.

3.1. Low versus high frequency updating

When a time series panel of mixed frequency is analyzed, one may consider the highest frequency to formulate the stationary

Table 2
Monte Carlo results II and III. Results II: For the same Monte Carlo study as carried out for Table 1, we report the frequencies of optimal W values in 500 Monte Carlo simulations. The optimal W is this value of W that produces the smallest MSE within each simulation; the WML method is used for the estimation of the parameters in the DFM from which the in-sample one-step ahead forecasts are computed. Results III: We report the finite sample realized rejection rate (RRR) of the Diebold–Mariano test in 500 Monte Carlo simulations based on the optimized value of W, see Results II. We adopt a 90% confidence level for the Diebold–Mariano test concerning forecasts obtained from WML against those from ML estimates. The results for correct specification can be considered as the *size* of the test and for 'misspecification' as the *power* of the test.

W	Correct specification			Misspecification		
	$N_x = 2$	$N_x = 5$	$N_x = 10$	$N_x = 2$	$N_x = 5$	$N_x = 10$
	II: Frequencies of optimal W values					
1	59.8	41.2	36.8	1.4	2.2	0.8
2	15	10.0	12.0	0.4	0.2	0.4
3	1.6	4.0	7.2	0.2	0.0	0.0
4	0.8	3.0	3.8	0.4	0.2	0.0
5	1.6	4.0	3.6	0.2	0.2	0.0
6	2.8	4.0	6.6	0.0	0.0	0.0
10	2.4	4.8	7.2	1.2	0.0	0.2
15	3.4	5.2	3.4	0.4	0.0	0.2
25	3.2	3.6	5.4	0.8	0.0	0.0
50	2.4	4.8	3.6	1.8	0.6	0.0
$\geq$100	7.0	15.4	10.4	93.2	96.6	98.4
	III: Rejection rates of Diebold–Mariano test					
RRR	0.044	0.084	0.118	0.914	0.928	0.934

dynamic process of the factor f_t in (1). For the low frequency variables, the entries corresponding to the high frequency time points, for which no observations are available, can be treated as missing observations. It is well established that the Kalman filter is designed to handle missing observations. This solution has been adopted in, among others, Mariano and Murasawa (2003) and Bańbura and Rünstler (2011), although these contributions consider different dynamic factor model specifications. An alternative solution is to specify the state vector of the underlying linear state space model as a low frequency variable. We show that the low frequency state vector can still accommodate the high frequency dynamics in the model, simultaneously with the low frequency dynamics. Also, in many cases of practical relevance, this solution can lead to a much lower computational burden. In the Online Appendix we show how monthly dynamics can be treated by a linear state space model with a low frequency time index, say a quarterly index, in which the monthly entries are stacked into a quarterly vector. Similar ideas can be adopted for even higher frequency dynamics for weekly and daily observation. The idea of stacking the series observed at higher frequencies into vectors of the lowest frequency has also been explored by Ghysels (2012) in a vector autoregressive context and by Marcellino et al. (2014) for explanatory variables in their Bayesian regression model with stochastic volatility. In our case we adopt these ideas for all high frequency variables in a dynamic factor model.

3.2. Low versus high frequency updating: computing times

To illustrate the computational consequences of low versus high frequency formulations, we consider a monthly autoregressive model of order 1, AR(1) model, for the variable x_τ^m, where τ is the high frequency time index and the superscript "m" is to indicate it is a monthly variable x, that is $x_\tau^m = \varphi x_{\tau-1}^m + \varepsilon_\tau^m$ where ε_τ^m is the corresponding zero-mean disturbance and φ is the autoregressive coefficient. We can stack three consecutive values of the monthly variable x_τ^m in a specific quarter t into the quarterly 3×1 vector x_t, where t is the low frequency time index. We have

$$x_t = \begin{pmatrix} x_{t,1} \\ x_{t,2} x_{t,3} \end{pmatrix} = \begin{pmatrix} x_\tau^m \\ x_{\tau+1}^m \\ x_{\tau+2}^m \end{pmatrix}. \quad (5)$$

where $x_{t,i}$ is the ith element of x_t and corresponds to the value of the x variable in month i of quarter t. The high frequency dynamic process for x_τ^m can be expressed by the low frequency vector process

$$\begin{pmatrix} x_{t,1} \\ x_{t,2} \\ x_{t,3} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \phi_x \\ 0 & 0 & \phi_x^2 \\ 0 & 0 & \phi_x^3 \end{pmatrix} \begin{pmatrix} x_{t-1,1} \\ x_{t-1,2} \\ x_{t-1,3} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ \phi_x & 1 & 0 \\ \phi_x^2 & \phi_x & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{t,1} \\ \varepsilon_{t,2} \\ \varepsilon_{t,3} \end{pmatrix}, \quad (6)$$

for $t = 1, \ldots, T$, where $(\varepsilon_{t,1}, \varepsilon_{t,2}, \varepsilon_{t,3})'$ is similarly defined as x_t in relation to the monthly disturbance ε_τ^m. The Online Appendix provides the derivation of this formulation and more detailed general results (see Appendix), such as for autoregressive models of higher orders p, that is AR(p) models. The dynamic properties of the monthly series x_τ^m are not altered by this quarterly formulation: the quarterly vector process is just a different formulation of the monthly process. Whether the monthly dynamic process is represented by x_τ^m or by the stacked quarterly 3×1 vector x_t, or even by a yearly 12×1 vector, it also 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 in the Online Appendix only rely on equalities. In all cases, the Kalman filter can be used for loglikelihood evaluation. Therefore the maximized loglikelihood value and the parameter estimates are the same for low and high frequency representations of autoregressive, or any other linear, dynamic processes.

However, the different representations have an effect on computing times. For example, for 100 years of data and for a monthly representation, we have a time series dimension of $T = 1200$. When the data is stacked into quarterly 3×1 vectors we have $T = 400$ and with yearly 12×1 vectors we only have $T = 100$. On the other hand, the stacked vector x_t will have a larger dimension. Hence the different representations will have an effect on the Kalman filter computations 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 1000 years, for different orders p and using different representations: daily, weekly, monthly, quarterly and yearly. For example, for a weekly AR(p) process, time series are generated consisting of $T = 52{,}000$ weekly observations. The loglikelihood value is then calculated 10,000 times using the parameter values that maximize the loglikelihood function, for different lower frequency representations. We have verified that likelihood evaluations for these different representations resulted in the same value.

The computing times for different combinations of AR(p) processes and frequencies are presented in Table 3. We focus here on the results for weekly and daily autoregressive processes; the results for monthly processes are presented in the Online Appendix. It is clear that for weekly and daily AR(1) and AR(2) processes, the representations based on weekly and daily models, respectively, are computationally 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 1000 iterations in the 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 stacked vectors are of the same size given the number of lags that need to be accommodated in the state vector. Similar effects take place when p increases further and even the yearly representation becomes computationally more efficient. The results for weekly and daily processes in Table 3 clearly illustrate that stacking observations into low frequency vectors can lead to large computational gains, especially when many lagged dependent variables are part of the model. The considered lag lengths may appear in some cases artificially large in Table 3, but we notice that a yearly AR(1) process can only be represented by a weekly AR(52) process.

Table 3
Computing times. The left panel of this table presents the total computing time that is required to filter 1000 weekly time series with $T = 52{,}000$ using Kalman filter for the corresponding weekly AR(p) model. Four different approaches are used: treating the data as weekly observations, stacking the data into 13 "monthly" 4×1 vectors, stacking the data into quarterly 13×1 vectors and stacking the data into yearly 52×1 vectors. The right panel of this table presents the total computing time that is required to filter 1000 daily time series with $T = 364{,}000$ using Kalman filter for the corresponding daily AR(p) model. Four different approaches are used: treating the data as daily observations, stacking the data into weekly 7 by 1 vectors, stacking the data into 13 "monthly" 28×1 vectors and stacking the data into quarterly 91×1 vectors. Each value presents the aggregate computing time over 1000 simulations. For each p, the shortest time of the four approaches is highlighted.

p	Weekly time series				p	Daily time series			
	Week	"Month"	Quarter	Year		Day	Week	"Month"	Quarter
1	2, 5	3, 5	6, 9	69, 9	1	2, 6	3, 9	15, 3	108, 4
2	3, 4	3, 8	7, 5	72, 5	2	3, 6	4, 2	16, 6	112, 7
3	4, 4	4, 0	8, 3	76, 1	3	4, 6	4, 5	17, 7	116, 4
6	14, 0	7, 8	10, 6	85, 4	6	14, 6	7, 0	21, 8	130, 7
7	18, 1	10, 1	11, 5	87, 3	7	18, 8	9, 2	29, 7	132, 0
8	24, 3	11, 9	12, 2	89, 9	8	24, 9	11, 5	24, 4	135, 3
11	47, 0	22, 4	14, 4	100, 2	18	379, 4	57, 2	36, 6	174, 5
12	54, 2	24, 9	15, 2	103, 2	19	287, 3	68, 5	39, 9	179, 0
13	65, 6	30, 1	18, 2	106, 0	20	319, 8	68, 3	38, 8	182, 2
46	6504, 6	609, 5	325, 3	209, 9	58	13333, 8	930, 8	435, 4	316, 3
47	8064, 1	647, 5	354, 2	209, 3	59	16124, 3	970, 4	455, 5	318, 6
48	9252, 8	684, 8	363, 3	220, 5	60	15937, 9	1005, 2	470, 9	323, 1

3.3. *Mixed frequency dynamic factor model: low frequency updating*

We provide the details of the low frequency representation of a dynamic factor model that contains both low and high frequency variables. To keep the discussion simple, and tailored to the DFM that we adopt for our empirical study in Section 4, we consider the DFM for the quarterly target variable y_t, with $N_y = 1$, and for a monthly $N_x \times 1$ vector x_τ^m, with quarterly index t and monthly index τ. We define the $(3 \cdot N_x) \times 1$ vector x_t in terms of x_τ^m as in (5).

The dynamic factor model (1) is formulated for a quarterly time index t. In case the autoregressive process of the dynamic factor is formulated as a monthly process, we introduce the monthly $r \times 1$ vector f_τ^m for month τ. The vector autoregressive process of order p^m is given by

$$f_\tau^m = \Phi_1^m f_{\tau-1}^m + \Phi_2^m f_{\tau-2}^m + \cdots + \Phi_p^m f_{\tau-p^m}^m + \eta_\tau^m,$$
$$\eta_\tau^m \sim N(0, \Sigma_\eta),$$

where Φ_j^m is the autoregressive coefficient matrix for lag j and η_τ^m is a monthly disturbance vector. The stack of three consecutive months of f_τ^m, corresponding to a specific quarter t, is f_t as in (5). In particular, we have $f_t = (f_{t,1}', f_{t,2}', f_{t,3}')' = (f_\tau^{m\prime}, f_{\tau+1}^{m\prime}, f_{\tau+2}^{m\prime})'$. The monthly dynamic processes of the elements of the stacked vector f_t can be formulated as a quarterly vector autoregressive process, similar to the equation for f_t in (1). The mixed frequency dynamic factor model is given by

$$\begin{pmatrix} y_t \\ x_{t,1} \\ x_{t,2} \\ x_{t,3} \end{pmatrix} = \begin{pmatrix} \Lambda_y & \Lambda_y & \Lambda_y \\ \Lambda_x & 0 & 0 \\ 0 & \Lambda_x & 0 \\ 0 & 0 & \Lambda_x \end{pmatrix} \begin{pmatrix} f_{t,1} \\ f_{t,2} \\ f_{t,3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t,1} \\ \varepsilon_{x,t,2} \\ \varepsilon_{x,t,3} \end{pmatrix} \tag{7}$$

where the $N_z \times r$ factor loading matrix $\Lambda = [\Lambda_y', \Lambda_x']'$ is used in (1) and where $\varepsilon_{x,t,j}$ is the disturbance associated with $x_{t,j}$ for month j in quarter t. The similarity with (1) becomes even clearer by formulating the model as

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{bmatrix} \iota_3' \otimes \Lambda_y \\ I_3 \otimes \Lambda_x \end{bmatrix} f_t + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix}, \tag{8}$$

where f_t is the $(3 \cdot r) \times 1$ vector of monthly factors and $\varepsilon_{x,t} = (\varepsilon_{x,t,1}', \varepsilon_{x,t,2}', \varepsilon_{x,t,3}')'$, with ι_k being the $k \times 1$ vector of ones. The dynamic specification for f_t, implied by any linear dynamic process for the monthly vector f_τ^m, can be formulated in state space form; see the Online Appendix. 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 $\Lambda_y = [\Lambda_{y,1}, \Lambda_{y,2}, \Lambda_{y,3}]$, where $\Lambda_{y,i}$ is the $N_y \times r$ factor loading matrix for month i. Other mixed frequencies than for monthly and quarterly variables can be considered. For example, a mix of yearly and weekly variables can be jointly modeled using the above stacking approach in a similar way.

Alternatively, we can specify the dynamic factor in our mixed frequency model as a quarterly process rather than a monthly process. In this case, the vector f_t is not a stacked vector; it just contains the r dynamic factors. For example, we can specify the dynamic process for f_t by the quarterly vector autoregressive model of order p, VAR(p),

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

with autoregressive coefficient matrix Φ_j, for lag $j = 1, \ldots, p$ in quarters. The observation equation of the dynamic factor model for quarterly factors is then simply given by

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{bmatrix} \Lambda_y \\ \iota_3 \otimes \Lambda_x \end{bmatrix} f_t + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix}. \tag{9}$$

In this case, we have f_t being applicable to all three months in quarter t.

4. Empirical study: forecasting U.S. GDP growth

We present the results of our empirical study where we investigate whether the WML method is able to improve the nowcasting and forecasting of a relevant economic time series, in our case quarterly growth in U.S. gross domestic product, by means of a mixed frequency dynamic factor model.

Table 4
Database variables. We present the labels and definitions of the quarterly, monthly and weekly variables that are used in our empirical study for the U.S. economy. All time series are available from 1970 up to 2009: forty years of data. The third column indicates the frequency of the series: monthly (M), quarterly (Q) or weekly (W). The last three columns indicate which variables are selected for the models (i)–(iii). Data is obtained from "FRED2".

Variable	Description	Frequency	Model selections		
	Target variable y_t		(i)	(ii)	(iii)
GDP	U.S. Real GDP (billions of chained 1996)	Q	✓	✓	✓
	Related variables x_t				
EMP	Employees on non-agricultural payrolls	M	✓	✓	✓
DPI	Real disposable personal income	M	✓	✓	✓
IPI	Industrial production index	M	✓	✓	✓
SLS	Manufacturing and trade sales	M	✓	✓	✓
PMI	Purchasing manager index, manufacturing	M		✓	
UR	Unemployment rate	M		✓	
PCE	Personal consumption expenditures	M		✓	
HS	Housing starts total	M		✓	
NRS	New residential sales	M		✓	
PPI	Producer price index, finished goods	M		✓	
MNO	Manufacturer new orders	M		✓	
CPI	Consumer price index, all urban consumers	M		✓	
PFS	Philadelphia Fed survey, business conditions	M		✓	
IUC	Initial unemployment claims	W			✓

4.1. Data

In our empirical analysis, we have constructed a database of 15 U.S. economic time series that are listed in Table 4. The target series is quarterly gross domestic product (GDP) growth. The time index t refers to a quarterly index. The other economic time series are observed at a monthly frequency except the series "initial unemployment claims" that is observed at a weekly frequency. The data is collected from the "FRED2" website. Our dataset can be regarded as a subset of the data used in the study by Bańbura et al. (2013). The time series length is 40 years, starting from 1970 and ending at 2009.

We consider three different panel selections from the database that are included in the monthly vector x_τ^m and the weekly variable $x_{\tau^*}^w$, where τ is the monthly index and τ^* is the weekly index. We assume that we have 52 of 53 weeks in one year; mostly we have 13 weeks in one quarter but occasionally we have 14 weeks in one quarter. Hence the dimension of x_t can vary in some occurrences. The three different selections are:

(i) The first four monthly related variables with labels EMP, DPI, IPI and SLS; see Table 4. This selection of variables is similar to the one considered by Mariano and Murasawa (2003) and Bańbura et al. (2013).
(ii) All monthly related variables listed in Table 4. This selection of 13 series is a subset of the 16 monthly variables used in the analysis of Bańbura et al. (2013).
(iii) The first four monthly related variables plus the last weekly variable IUC listed in Table 4. This selection mimics to some extent the analysis of Aruoba et al. (2009) where also quarterly, monthly and weekly variables are used.

All time series are transformed and/or demeaned so that no intercept coefficients are required in the model. Detected outliers in each series are replaced by their median values of the previous five observations; here we follow Stock and Watson (2005).

4.2. Design of the empirical study

In our empirical study, we investigate the forecasting performances of mixed frequency dynamic factor models (*mf* DFMs) with their parameters estimated by standard maximum likelihood (ML) and by our weighted maximum likelihood (WML) method. We aim to establish whether the WML method for different *mf* DFMs can improve the accuracy of nowcasts and forecasts of quarterly growth in U.S. gross domestic product (quarterly GDP growth). Our study is a stylized exercise to investigate the role of the WML method in empirical settings; we do not present results of a real-time forecasting study.

We consider three different *mf* DFM specifications corresponding to the three database selections described above. The *mf* DFM is formulated in terms of the quarterly time index t for y_t and x_t where x_t consists of stacks of three monthly observations (and, possibly, of a stack of thirteen, occasionally fourteen, weekly observations), which correspond to quarter t, for each related variable in the selected panel. Apart from the three panel selections, we also consider two variants of which the first variant formulates the dynamic factor as a monthly variable (M) as in (8) and the second has the dynamic factor formulated as a quarterly variable (Q) as in (9). The label *mf* DFM[j, q] refers to the *mf* DFM specification for panel j and frequency q, with j = (i), (ii), (iii) and $q = Q, M$. In all cases, the vector autoregressive processes for the dynamic factors are based on one lag as in (1).

In our empirical study we assess the improvements in the forecasting and nowcasting accuracy by adopting the WML method for different weights W. The forecast evaluation period for the quarterly target time series y_t starts from 2000, quarter 1, until 2009, quarter 4, that is forty quarterly observations. For each forecast, the earlier thirty year sample is used for parameter estimation and determining the W in our WML method. The first twenty years are used for parameter estimation and the last ten years are used for determining the optimal weights, within a rolling sample scheme; see the discussion in Section 2.3. Then the nowcasts, the h-steps ahead (measured in months) forecasts and their subsequent errors are computed, for $h = 1, 2, 3, 6, 12$. The details are as follows. The forecasts are calculated using a rolling window of ten years of data to estimate the parameters and determine the weights. We evaluate both the nowcasting ($h = 1, 2$ months) and the forecasting ($h = 3, 6, 12$ months) performance of all the competing models. When $h = 1$, the values of x_t are known until the first two months of the quarter that needs to be forecasted. When $h = 2$, only the first month of the quarter to be forecasted 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 $h = 6$ and $h = 12$ we are forecasting two and four quarters ahead, respectively.

Table 5
Forecasting comparisons for the quarterly U.S. GDP growth rate between 2000 and 2009 at forecast horizons $h = 1, 2, 3, 6, 12$ months (a 10 years rolling window). We present the mean squared forecast errors (MSEs) for the mfDFM(j, q) models with parameters estimated by maximum likelihood (weight $W = 1$) and by weighted maximum likelihood (weight $W = W^{opt}$). The optimal weight W^{opt} is determined as described in Section 4.2. For each forecasting horizon the most accurate model (with smallest MSE) is highlighted.

j	q	W	$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$
(i)	M	1	0.8138	0.8139	0.8138	0.8138	0.8138
		W^{opt}	0.5493	0.8138	0.6580	0.7715	0.8100
	Q	1	0.6399	0.6654	0.7258	0.8040	0.8199
		W^{opt}	0.6084	0.8138	0.6949	0.7859	0.8173
(ii)	M	1	0.5306	0.8138	0.8138	0.8138	0.8619
		W^{opt}	0.5318	0.5760	0.6184	0.7407	0.8138
	Q	1	0.5353	0.5905	0.6686	0.7842	0.8138
		W^{opt}	0.5259	0.5597	0.6345	0.7813	0.8539
(iii)	W	1	0.7848	0.7847	0.7847	0.7847	0.7847
		W^{opt}	0.5438	0.5917	0.7847	0.7847	0.7847
	Q	1	0.7848	0.7848	0.6768	0.7848	0.8149
		W^{opt}	0.5255	0.5735	0.7847	0.7677	0.8082

Table 6
Relative forecasting comparisons for the quarterly U.S. GDP growth rate between 2000 and 2009 at forecast horizons $h = 1, 2, 3, 6, 12$ months (a 10 years rolling window). We present the MSE ratios of forecasts based on WML estimates relative to forecasts based on ML estimates, considering the same mfDFM(j, q) model specifications. The values in parentheses are the p-values of the corresponding Diebold–Mariano test; its significance at the 10% confidence level is indicated by an asterix.

j	q	$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$
(i)	M	0.6749*	0.9999	0.8086	0.9480	0.9953
		(0.099)	(0.191)	(0.138)	(0.168)	(0.416)
	Q	0.9508	1.2231	0.9574	0.9775	0.9968
		(0.171)	(0.802)	(0.138)	(0.193)	(0.336)
(ii)	M	1.0023	0.7079	0.7599	0.9102	0.9442
		(0.519)	(0.103)	(0.142)	(0.213)	(0.144)
	Q	0.9826	0.9479*	0.9491*	0.9962	1.0494
		(0.344)	(0.057)	(0.094)	(0.412)	(0.869)
(iii)	W	0.6930*	0.7540	0.9999	0.9999	0.9999
		(0.052)	(0.140)	(0.108)	(0.126)	(0.136)
	Q	0.6696*	0.7308	1.1595	0.9782	0.9918
		(0.038)	(0.116)	(0.757)	(0.356)	(0.238)

We evaluate the accuracy of the forecasting and nowcasting of quarterly GDP growth by means of the mean squared error (MSE) based on the last forty quarters (out-of-sample). In case of nowcasting ($h = 1, 2$) we have 40 errors, in case of one-quarter ahead forecasting ($h = 3$) we have 39 errors, etc. We compare the resulting MSEs but we also assess whether the nowcasts and forecasts based on WML estimates outperform those based on ML estimates by means of the Diebold–Mariano (DM) test.

4.3. Empirical results: forecast performance

The forecasting results for our mixed frequency dynamic factor models with different selections of related variables, with different dynamics for the dynamic factors, and with parameters estimated by two different methods are presented in Table 5. The forecast accuracy is measured by the mean squared forecasting errors (MSE). In the case of the WML estimation method, we take different integer values for W and determine its optimal value over a large range. The selected optimal value is denoted by W^{opt} and is used for the forecasting exercise. The most accurate model for each forecasting horizon is determined by the smallest MSE. In particular we focus on comparing the forecasting accuracy between ML and WML methods. Some of the differences reported in Table 5 are large. For example, in the case of model mfDFM[(i), M] and one-step ahead forecasting, $h = 1$, the MSE drops from 0.81 when parameters are estimated by ML, to 0.55 when estimated by WML. In almost all cases, amongst all forecast horizons, the WML method leads to a smaller MSE when the same model is considered but with parameters estimated by ML. However, differences in MSEs also occur when we compare amongst different models and different frequencies in which the dynamics are modeled. We may conclude from Table 5 that the most accurate nowcasts ($h = 1, 2$) are obtained from the WML method applied to the larger panel model (ii) with quarterly factors. The most accurate forecasts ($h > 2$) are also obtained from WML but with monthly or weekly factors.

In earlier studies, amongst others, Aruoba et al. (2009) and Bańbura et al. (2013) argue that incorporating weekly data into a mixed frequency dynamic factor model is beneficial to nowcasting. Our results are clearly consistent with these findings. The improvements delivered by model (iii), in comparison to model (i), and especially when parameters are estimated by WML, are substantial in all cases. We should notice that model (iii) is the same as model (i) with only the weekly variable "initial unemployment claims" added to it. The improvements are less convincing for forecasting. In case of within the year forecasting ($h = 3, 6$), model (ii) produces the smallest MSE overall. In case of one-year ahead forecasting ($h = 12$), model (iii) is preferred overall.

4.4. Empirical results: relative forecast performance

The relative MSE ratios between the two estimation methods are presented in Table 6. The entries are relative to the MSE from the ML method: a value smaller than 1 implies that the WML method provides more accurate forecasts (in terms of a smaller MSE) when compared to the MSE from the ML method, based on the same model specification for mfDFM[j, q]. We may conclude from the results presented in Tables 5 and 6 that the WML method improves the forecasting accuracy in most cases. The MSE reductions are even large when the ML method provides relative inaccurate forecasts. For example, the nowcasting MSE value for $h = 1$ and model mfDFM[(iii), Q], with estimated parameters by the ML method, is 0.78 which is a relatively large value amongst the MSE values for the other competing models. But when the parameters are estimated by our WML method, the MSE is reduced to 0.53, which is the smallest value amongst all other models. The forecasting accuracy is improved by about 33% and this improvement is significant at the 10% confidence level. Such results are consistent with the argument in Section 2. When the model is misspecified, achieving a good in-sample fit does not lead to a good out-of-sample forecasting accuracy. Our WML method is designed to perform well when the model is misspecified and interest lies in forecasting only the target variables.

When concentrating on models with quarterly factors ($q = Q$), we may conclude from the reported MSEs in Table 5 as follows. Amongst the three mixed frequency DFMs, as indicated in Table 4, model (ii) is superior in nowcasting ($h = 1, 2$) compared to models (i) and (iii), when parameters are estimated by the ML method. In case of forecasting ($h > 2$), the performances of the models (ii) and (iii) are similar overall but outperform model (i), when the ML method is used. When we focus on these comparisons for the WML method, our findings change somewhat. For both nowcasting and forecasting, the models (ii) and (iii) perform equally well overall but clearly outperform model (i).

We adopt the Diebold–Mariano (DM) test to verify whether the forecasts obtained from the WML method are significantly more accurate compared to those obtained from the ML method, when considering the same model. The p value of the DM test for each model is also presented in Table 6. Almost all MSE ratios are

Table 7
Relative forecasting comparisons for quarterly U.S. GDP growth rate, for periods (a) between 2000 and 2007, and (b) 2008 and 2009. For further details, see Table 6.

		(a) Period 2000–2007					(b) Period 2008 and 2009				
		$h=1$	$h=2$	$h=3$	$h=6$	$h=12$	$h=1$	$h=2$	$h=3$	$h=6$	$h=12$
(i)	M	0.8245*	0.9999	1.0023	0.9407	1.0326	0.5966*	0.9999	0.7058*	0.9512*	0.9783*
		(0.054)	(0.177)	(0.510)	(0.327)	(0.815)	(0.068)	(0.154)	(0.049)	(0.006)	(0.095)
	Q	0.8925*	0.9971	0.9576*	0.9840	0.9924	0.9995	1.3885	0.9572	0.9744	0.9990
		(0.035)	(0.492)	(0.065)	(0.222)	(0.186)	(0.497)	(0.811)	(0.288)	(0.226)	(0.467)
(ii)	M	0.9859	0.8912	1.0077	1.0525	0.9580*	1.0259	0.6090*	0.6189*	0.8393*	0.9383
		(0.340)	(0.135)	(0.541)	(0.660)	(0.000)	(0.598)	(0.053)	(0.013)	(0.045)	(0.288)
	Q	0.9632	0.9374*	0.9735	0.9926	1.0578	1.0089	0.9574	0.9257	0.9981	1.0456
		(0.231)	(0.056)	(0.101)	(0.165)	(0.967)	(0.546)	(0.206)	(0.112)	(0.480)	(0.668)
(iii)	W	0.9567	0.9595	0.9999	0.9999	1.0000	0.5367*	0.6426	0.9999	0.9999	0.9999
		(0.345)	(0.346)	(0.123)	(0.365)	(0.579)	(0.052)	(0.118)	(0.136)	(0.176)	(0.187)
	Q	0.9133	0.9107	0.9356	1.0410	0.9985	0.5273*	0.6353	1.3273	0.9492*	0.9885
		(0.190)	(0.162)	(0.232)	(0.639)	(0.465)	(0.044)	(0.108)	(0.856)	(0.003)	(0.221)

smaller than unity for the different forecasting horizons, but for a number of entries we have forecasts from the WML method that are significantly better than those from the ML method, at the 10% confidence level. In particular, the WML method has significantly improved the forecasting accuracies for models (i) and (iii) with high frequency factors M and W, for the nowcasting horizon $h = 1$.

To obtain a more detailed picture of the empirical relevance of our WML method, we also study the improvements in forecasting accuracy for two different forecasting periods: (a) from 2000 to 2007, and (b) from 2008 to 2009. The second forecasting period (b) covers the enduring period of the financial crisis. The relative MSE ratios for the WML method with respect to the ML method are presented in Table 7, in the same way as in Table 6. The most significant forecasting accuracy improvements by the WML method are realized in the financial crisis period (b). We already have concluded that the WML method is expected to improve forecast accuracy when models are misspecified. Given that after the financial crisis the forecasting ability of most models have deteriorated, the WML method can be quite effective in reducing the nowcast and forecast MSEs. It can also be concluded from the results presented in Table 7 that overall the WML method is most effective for short-term forecasting or nowcasting ($h = 1, 2$). At forecasting horizon $h = 1$, many of the reported WML improvements are high or present a significant Diebold–Mariano test statistic, especially for models (i) and (iii).

5. Conclusions

We have introduced a new weighted maximum likelihood (WML) estimation procedure for dynamic factor models where some target variables are of key importance while the other, related, variables are only used to facilitate the forecasting or analysis of target variables. The WML method introduces variable-specific weights in the likelihood function to let the target variables have more importance than the related variables in the parameter estimation process. We have derived the asymptotic properties of the WML estimator of the parameter vector and have provided an information-theoretic characterization based on the Kullback–Leibler divergence. Furthermore, we have proposed a cross-validation method to estimate the weights that optimize the forecasting performance for the target variable. The Monte Carlo study for investigating the finite sample performance of the WML estimator highlights its good properties overall. In empirical studies, the dynamic factor model may need to be modified to accommodate a panel of time series in which variables are observed at different seasonal frequencies such as monthly, quarterly and weekly. We have argued that a low frequency representation of the mixed frequency dynamic factor model leads to a computationally efficient treatment. The representation is highly flexible and can allow for low and high frequency dynamic factors. The empirical study shows that our WML method can lead to significant improvements in the accuracy of nowcasting and forecasting U.S. GDP growth. We expect that our proposed solutions also have consequences in other applications and in other modeling frameworks. Interesting future research may focus on applying the WML method to other models and settings as well as providing a more detailed comparison with other estimation approaches in which selections of variable equations can be penalized. For example, a possible alternative approach is to adopt Bayesian estimation methods where different prior conditions are given to parameters that are associated with different variable equations.

Acknowledgment

S.J. Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series, at Aarhus University, funded by the Danish National Research Foundation (DNRF78).

Appendix. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jeconom.2016.04.014.

References

Aruoba, S.B., Diebold, F.X., Scotti, C., 2009. Real-time measurement of business conditions. J. Bus. Econom. Statist. 27 (4), 417–427.
Bańbura, M., Giannone, D., Modugno, M., Reichlin, L., 2013. Now-casting and the real-time data flow. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 2. Elsevier, pp. 195–237.
Bańbura, M., Modugno, M., 2014. Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. J. Appl. Econometrics 29 (1), 133–160.
Bańbura, M., Rünstler, G., 2011. A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. Int. J. Forecast. 27 (2), 333–346.
Bittanti, S., Colaneri, P., 2000. Invariant representations of discrete-time periodic systems. Automatica 36 (12), 1777–1793.
Bittanti, S., Colaneri, P., 2009. Periodic Systems: Filtering and Control. Springer-Verlag.
Blasques, F., Koopman, S.J., Lucas, A., 2014. Maximum Likelihood Estimation for Generalized Autoregressive Score Models. Technical Report 14-029/III, Tinbergen Institute Discussion Paper.
Bougerol, P., 1993. Kalman filtering with random coefficients and contractions. SIAM J. Control Optim. 31 (4), 942–959.
Bougerol, P., Picard, N., 1992. Strict stationarity of generalized autoregressive processes. Ann. Probab. 20, 1714–1730.
Bräuning, F., Koopman, S.J., 2014. Forecasting macroeconomic variables using collapsed dynamic factor analysis. Int. J. Forecast. 30 (3), 572–584.
Chen, W., Anderson, B.D., Deistler, M., Filler, A., 2012. Properties of blocked linear systems. Automatica 48 (10), 2520–2525.

Clark, T.E., McCracken, M.W., 2001. Tests of equal forecast accuracy and encompassing for nested models. J. Econometrics 105 (1), 85–110.

Clark, T.E., McCracken, M.W., 2015. Nested forecast model comparisons: A new approach to testing equal accuracy. J. Econometrics 186 (1), 160–177.

Diebold, F.X., 2012. Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests. NBER Working Papers 18391, National Bureau of Economic Research, Inc.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econom. Statist. 13 (3), 253–263.

Doz, C., Giannone, D., Reichlin, L., 2011. A two-step estimator for large approximate dynamic factor models based on Kalman filtering. J. Econometrics 164 (1), 188–205.

Durbin, J., Koopman, S.J., 2012. Time Series Analysis by State Space Methods. Oxford University Press, Oxford.

Eguchi, S., Copas, J., 1998. A class of local likelihood methods and near-parametric asymptotics. J. R. Stat. Soc. Ser. B Stat. Methodol. 60 (4), 709–724.

Engle, R.F., Watson, M.W., 1981. A one-factor multivariate time series model of metropolitan wage rates. J. Amer. Statist. Assoc. 76, 774–781.

Foroni, C., Gurin, P., Marcellino, M., 2015. Markov-switching mixed-frequency VAR models. Int. J. Forecast. 31 (3), 692–711.

Ghysels, E., 2012. Macroeconomics and the Reality of Mixed Frequency Data.

Ghysels, E., Hill, J.B., Motegi, K., 2015. Testing for Granger causality with mixed frequency data. J. Econometrics (forthcoming).

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74 (6), 1545–1578.

Giannone, D., Reichlin, L., Small, D., 2008. Nowcasting: the real-time informational content of macroeconomic data. J. Monet. Econ. 55 (4), 665–676.

Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.

Hu, F., Zidek, J., 1997. The relevance weighted likelihood (unpublished).

Jungbacker, B., Koopman, S.J., 2015. Likelihood-based dynamic factor analysis for measurement and forecasting. Econom. J. 18, C1–C21.

Koopman, S.J., Durbin, J., 2000. Fast filtering and smoothing for multivariate state space models. J. Time Series Anal. 21, 281–296.

Marcellino, M., Carriero, A., Clark, T.E., 2014. Real-Time Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility. Working Paper.

Marcellino, M., Stock, J.H., Watson, M., 2003. Macroeconomic forecasting in the euro area: country specific versus euro wide information. Eur. Econom. Rev. 47, 1–18.

Mariano, R., Murasawa, Y., 2003. A new coincident index of business cycles based on monthly and quarterly series. J. Appl. Econometrics 18, 427–443.

Markatou, M., Basu, A., Lindsay, B.G., 1997. Weighted likelihood estimating equations: The discrete case with applications to logistic regression. J. Statist. Plann. Inference 57 (2), 215–232.

Markatou, M., Basu, A., Lindsay, B.G., 1998. Weighted likelihood equations with bootstrap root search. J. Amer. Statist. Assoc. 93 (442), 740–750.

Mehra, R., 1970. On the identification of variances and adaptive Kalman filtering. IEEE Trans. Automat. Control 15 (2), 175–184.

Mittnik, S., Zadrozny, P.A., 2005. Forecasting quarterly german gdp at monthly intervals using monthly ifo business conditions data. In: Sturm, J.-E., Wollmershauser, T. (Eds.), Ifo survey Data in Business Cycle and Monetary Policy Analysis. Physica-Verlag, Heidelberg, pp. 19–48.

Quah, D., Sargent, T.J., 1993. A dynamic index model for large cross sections. In: Stock, J.H., Watson, M. (Eds.), Business Cycles, Indicators and Forecasting. University of Chicago Press, Chicago, pp. 285–306.

Rünstler, G., Barhoumi, K., Benk, S., Cristadoro, R., den Reijer, A., Jakaitiene, A., Jelonek, P., Rua, A., Ruth, K., Nieuwenhuyze, C.V., 2009. Short-term forecasting of GDP using large datasets: a pseudo real-time forecast evaluation exercise. J. Forecast. 28 (7), 595–611.

Schumacher, C., Breitung, J., 2008. Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. Int. J. Forecast. 24 (3), 386–398.

Staniswalis, J.G., 1989. The kernel estimate of a regression function in likelihood-based models. J. Amer. Statist. Assoc. 84 (405), 276–283.

Stock, J.H., Watson, M., 2002a. Forecasting using principal components from a large number of predictors. J. Amer. Statist. Assoc. 97, 1167–1179.

Stock, J.H., Watson, M., 2002b. Macroeconomic forecasting using diffusion indexes. J. Bus. Econom. Statist. 20, 147–162.

Stock, J.H., Watson, M., 2005. Implications of dynamic factor models for VAR analysis. Discussion Paper.

Straumann, D., Mikosch, T., 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equation approach. Ann. Statist. 34, 2449–2495.

Tibshirani, R., Hastie, T., 1987. Local likelihood estimation. J. Amer. Statist. Assoc. 82 (398), 559–567.

Wang, X., van Eeden, C., Zidek, J.V., 2004. Asymptotic properties of maximum weighted likelihood estimators. J. Statist. Plann. Inference 119 (1), 37–54.

Wang, X., Zidek, J.V., 2005. Selecting likelihood weights by cross-validation. Ann. Statist. 33 (2), 463–500.

Watson, M.W., Engle, R.F., 1983. Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression.. J. Econometrics 23, 385–400.

West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica 64 (5), 1067–1084.