TI 2014-105/III
Tinbergen Institute Discussion Paper

Low Frequency and Weighted Likelihood Solutions for Mixed Frequency Dynamic Factor Models

*Francisco Blasques**
*Siem Jan Koopman**
Max Mallee

* *Tinbergen Institute*

Faculty of Economics and Business Administration, VU University Amsterdam.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Low Frequency and Weighted Likelihood Solutions for Mixed Frequency Dynamic Factor Models

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

August 10, 2014

Abstract

The multivariate analysis of a panel of economic and financial time series with mixed frequencies is a challenging problem. The standard solution is to analyze the mix of monthly and quarterly time series jointly by means of a multivariate dynamic model with a monthly time index: artificial missing values are inserted for the intermediate months of the quarterly time series. In this paper we explore an alternative solution for a class of dynamic factor models that is specified by means of a low frequency quarterly time index. We show that there is no need to introduce artificial missing values while the high frequency (monthly) information is preserved and can still be analyzed. We also provide evidence that the analysis based on a low frequency specification can be carried out in a computationally more efficient way. A comparison study with existing mixed frequency procedures is presented and discussed. Furthermore, we modify the method of maximum likelihood in the context of a dynamic factor model. We introduce variable-specific weights in the likelihood function to let some variable equations be of more importance during the estimation process. We derive the asymptotic properties of the weighted maximum likelihood estimator and we show that the estimator is consistent and asymptotically normal. We also verify the weighted estimation method in a Monte Carlo study to investigate the effect of different choices for the weights in different scenarios. Finally, we empirically illustrate the new developments for the extraction of a coincident economic indicator from a small panel of mixed frequency economic time series.

Keywords: Asymptotic theory, Forecasting, Kalman filter, Nowcasting, State space.
JEL classification: C13, C32, C53, E17.

1 Introduction

The multivariate analysis of a panel of economic and financial time series with mixed frequencies can be treated by a range of different approaches. The problem of mixed frequency time series is regarded as a challenging problem in many applied econometric

*Contact: f.blasques@vu.nl, s.j.koopman@vu.nl and m.i.p.mallee@vu.nl

studies. Currently there are two main competing approaches in the literature to handle mixed frequency time series: *partial model* and *full system* methods. This classification is adopted from Banbura, Giannone, Modugno, and Reichlin (2013). For the partial model solution, the multivariate model specifications focus particularly on low frequency time series variables while high frequency explanatory variables are aggregated to the lower frequency. The dynamics of the explanatory variables are not described by the model. When the full system method is adopted, the low and high frequency variables are modeled simultaneously. This approach can model feedback between variables since all variables are endogenous. In most full system methods all variables are modeled at the high frequency where series observed at a low frequency have missing values.

The most commonly used partial model methods are Bridge models and Mixed Data Sampling (MIDAS) models. In Bridge models the high frequency data are forecasted up to the desired forecast horizon in a separate time series model. These forecasts are then aggregated to the lower frequency and are used as explanatory variables in a lower frequency time series model as contemporaneous values. Bridge models are often used to forecast quarterly Gross Domestic Product (GDP) using a set of monthly observed indicators. Trehan (1989) is the first application of Bridge equations in this setting. Baffigi, Golinelli, and Parigi (2004) and Golinelli and Parigi (2007) use Bridge models where GDP is predicted by the National Accounts income-expenditure identity. The MIDAS approach was proposed by Ghysels, Santa-Clara, and Valkanov (2006). As in Bridge models, the series with the lower frequency are regressed on the series with the higher frequency. However, when forecasting with MIDAS regression only one step is required. To obtain forecasts at horizon h, the values of y_t are simply regressed on the values of the indicators up to period $t - h$ and the dynamics of the regressors are not specified by the model. In MIDAS regressions, the higher frequency series are not aggregated, but each lag has its own regression coefficient. To avoid parameter proliferation, the coefficients of the different lags are described by a weighting function. Foroni, Marcellino, and Schumacher (2012) propose the use of unconstrained distributed lags of the high frequency indicators. This approach is referred to as unrestricted MIDAS or U-MIDAS.

Marcellino, Carriero, and Clark (2014) propose a method for producing current-quarter forecasts of GDP growth with a large range of available within-the-quarter monthly observations of economic indicators. Each time series of monthly indicators is transformed into three quarterly time series, each containing observations for, respectively, the first, second or third month of the quarter. Hence, there can be missing observations at the end of some of these three time series, depending on the specific month of the quarter. They include in the model only the constructed quarterly series without missing observations at the moment in time the forecast is formed. They use Bayesian methods to estimate the resulting model, which expands in size as more monthly data on the quarter become available.

Popular examples of full system methods are common factor models and mixed frequency vector autoregressive (VAR) models. Common factor models are a way to exploit

the whole available information set by condensing all time series into a few factors. One popular example is the construction of composite indicators (CI). These can be constructed for forecasting or for describing the current state of the economy. Stock and Watson (1990) construct a coincident index by applying maximum likelihood analysis to four monthly coincident indicators. Mariano and Murasawa (2003) extend this model by including quarterly GDP data and thus allowing for mixed frequency data. The model is cast into state space form and the likelihood is maximized using the Kalman Filter. In their model, all variables are driven by one unobserved monthly common factor and by an idiosyncratic factor for each individual series.

In the approach of Mariano and Murasawa (2003) the model operates at the highest frequency in the data, so all series are treated as monthly time series. All variables are assumed to be generated, but not necessarily observed, at this highest frequency, and thus can be used to produce forecasts of any variable at this frequency. The quarterly variable (GDP growth) is observed every third month and has missing values for the first two months of every quarter. This series is interpolated in order to estimate the unobserved latent monthly GDP. The suggested filtering algorithm is only an approximation. Aruoba, Diebold, and Scotti (2008) extend this model, but avoid approximations. They employ a one factor dynamic model to extract the unobserved state of the US economy, using four variables, including real high frequency (weekly and daily) data. The inclusion of high frequency daily data does not really change the picture compared with using only monthly indicators, but updates are available sooner.

Wohlrabe (2009) provides an extensive and detailed overview of the different models that have been explored using both of these approaches. Extensive descriptions of all model specifications and many empirical examples with comparisons are provided.

In this paper we propose a new method, in which the higher frequency data is stacked into a vector of observations which has the lower frequency. This is similar to the method proposed by Marcellino, Carriero, and Clark (2014). However, we propose a full system approach, in which the dynamics of all variables are described by the model. To discuss our methods in more specific terms, we consider a dynamic factor model, similar to the one proposed by Mariano and Murasawa (2003). We show that there is no need to introduce artificial missing values in the analysis while all the high frequency information is preserved and can still be analyzed in a computationally feasible way. Also the high frequency (monthly) optimal updating of new information can remain without the need to increase computational complexities.

Furthermore, we modify the method of maximum likelihood for parameter estimation by giving extra weight to one or more specific variables in a full system model approach. For example, when a dynamic factor model is adopted for the nowcasting and forecasting of a quarterly growth in gross domestic product (GDP), more weight can be given to GDP growth in comparison to the other variables in the dynamic factor model. For this purpose we introduce variable-specific weights in the likelihood function. We first discuss the asymptotic properties of this weighted maximum likelihood estimator and we

show that the estimator is consistent and asymptotically normal. We further verify our new approach in a Monte Carlo study to investigate the effect of different choices for the weights in different scenarios. We also adopt the weighted likelihood function for parameter estimation in our empirical study concerning the nowcasting and forecasting of US GDP growth based on a full system dynamic factor model with mixed frequency variables.

The outline of the paper is as follows. In Section 2 we show how high frequency autoregressive models can be specified as observationally equivalent models with a lower frequency. This method is illustrated by means of univariate monthly autoregressive (AR) processes, which we then formulate as quarterly and yearly AR processes. In many cases these new formulations lead to computational gains. In Section 3 we show how these transformations can be used to simultaneously model variables with different frequencies. We describe the specifications of our new approach in detail and also describe the models of other approaches, which will be used as benchmark models in the empirical section. In Section 4 we present our weighted maximum likelihood approach that is used for the focusing on key variables within a full system model. Asymptotic properties of the resulting estimator are derived. We also explore its small-sample properties in a Monte Carlo study and discuss the effect of using different weights in different scenarios. In Section 5 we present and explore the results of our empirical study concerning US GDP growth. We compare the nowcasting and forecasting accuracies of our new approach with those of benchmark aggregation and interpolation methods. We also study the empirical relevance of weighted estimation. Section 6 summarizes and concludes.

2 Low Frequency Solution for Univariate Models

2.1 Notation

In this section we discuss a stacking approach for analyzing high frequency univariate observations in a low frequency vector model. We first consider the case of monthly data (high frequency) that is represented by quarterly (low frequency) vectors. We use the notation x_τ^m for a variable x that is observed on a monthly (m) basis with monthly time index τ. The observations of the time series x_τ^m can be stacked into a quarterly (q) observed 3×1 vector x_t^q with quarterly time index t

$$x_t^q = \begin{pmatrix} x_{t,1}^q \\ x_{t,2}^q \\ x_{t,3}^q \end{pmatrix} = \begin{pmatrix} x_{3(t-1)+1}^m \\ x_{3(t-1)+2}^m \\ x_{3(t-1)+3}^m \end{pmatrix}, \tag{1}$$

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

In a similar way, we can represent the monthly observations into yearly vectors. The

monthly observed series x_τ^m can be stacked into a yearly (y) observed 12×1 vector x_t^y with yearly time index t. We then have

$$x_t^y = \begin{pmatrix} x_{t,1}^y \\ x_{t,2}^y \\ \vdots \\ x_{t,12}^y \end{pmatrix} = \begin{pmatrix} x_{12(t-1)+1}^m \\ x_{12(t-1)+2}^m \\ \vdots \\ x_{12(t-1)+12}^m \end{pmatrix}, \tag{2}$$

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

2.2 Linear State Space Models

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

$$\begin{aligned} x_t &= Z\alpha_t + \epsilon_t, & \epsilon_t &\sim NID\,(0, H)\,, \\ \alpha_{t+1} &= T\alpha_t + R\eta_t, & \eta_t &\sim NID\,(0, Q)\,, \end{aligned} \tag{3}$$

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

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

The initial state vector α_1 is generated from $\mathrm{N}(a_1,\ P_1)$, independently of $\epsilon_1, \ldots, \epsilon_n$ and $\eta_1, \ldots, \eta_n$, where a_1 and P_1 are assumed known, although P_1 may depend on the parameter vector ψ.

2.3 Autoregressive Processes

Autoregressive structures are the key ingredient for many dynamic linear models and therefore we will adopt autoregressive models to illustrate our stacking approach. We adopt our stacking method for other dynamic models in Section 3 as well.

Suppose that the univariate monthly observed variable x_τ^m is modeled by the autoregressive (AR) process with p lagged dependent variables, that is

$$x_\tau^m = \phi_1 x_{\tau-1}^m + \phi_2 x_{\tau-2}^m + \ldots + \phi_p x_{\tau-p}^m + \varepsilon_\tau^m, \qquad \varepsilon_\tau^m \sim NID\left(0, \sigma_\varepsilon^2\right), \tag{4}$$

where $\phi_1, \ldots, \phi_p$ are fixed and unknown autoregressive coefficients for monthly lags and ε_τ^m represents a serially uncorrelated Gaussian disturbance sequence with zero mean and a fixed and unknown variance σ_ε^2.

In the remainder of this subsection we provide quarterly model specifications for monthly AR processes. Some derivations and the initial properties of the state vector are provided in Appendix A. The generalizations to yearly, or other low frequencies, model specifications are straightforward.

AR(1) process

The monthly observations from the AR(1) process $x_\tau^m = \phi x_{\tau-1}^m + \varepsilon_\tau^m$ are stacked into the quarterly 3×1 vector x_t^q of (1). The quarterly process of the stacked variable x_t^q is then given by the vector autoregressive process

$$x_t^q = T x_{t-1}^q + R\varepsilon_t^q \tag{5}$$

with

$$T = \begin{pmatrix} 0 & 0 & \phi \\ 0 & 0 & \phi^2 \\ 0 & 0 & \phi^3 \end{pmatrix}, \qquad R = \begin{pmatrix} 1 & 0 & 0 \\ \phi & 1 & 0 \\ \phi^2 & \phi & 1 \end{pmatrix}, \tag{6}$$

such that the variance matrix of the vector x_t^q, conditional on x_{t-1}^q, is equal to $\sigma_\varepsilon^2 RR'$. We notice that all three elements of x_t^q depend only on the last element of x_{t-1}^q and on the associating elements of the vector of disturbances ε_t^q. The vector ε_t^q is the result of stacking the values of ε_τ^m in similar fashion as in (1). The autoregressive process (5) is equal to the linear Gaussian state space model (3) with state vector $\alpha_t = x_t^q$ and with system matrices T and R given by (6) and $Z = I_3$, $H = 0$, $Q = \sigma_\varepsilon^2$ and $\eta_t = \varepsilon_t^q$.

AR(3) process

The monthly observations from the AR(3) process $x_\tau^m = \phi_1 x_{\tau-1}^m + \phi_2 x_{\tau-2}^m + \phi_3 x_{\tau-3}^m + \varepsilon_\tau^m$ are stacked into the quarterly 3×1 vector x_t^q of (1). The quarterly process of the stacked variable x_t^q is again given by the vector autoregressive process (5) with

$$T = \begin{pmatrix} \phi_3 & \phi_2 & \phi_1 \\ \phi_1\phi_3 & \phi_1\phi_2 + \phi_3 & \phi_1^2 + \phi_2 \\ \phi_1^2\phi_3 + \phi_2\phi_3 & \phi_1^2\phi_2 + \phi_1\phi_3 + \phi_2^2 & \phi_1^3 + 2\phi_1\phi_2 + \phi_3 \end{pmatrix},$$
$$R = \begin{pmatrix} 1 & 0 & 0 \\ \phi_1 & 1 & 0 \\ \phi_1^2 + \phi_2 & \phi_1 & 1 \end{pmatrix}, \tag{7}$$

The variance matrix of the vector x_t^q, conditional on x_{t-1}^q, is equal to $\sigma_\varepsilon^2 RR'$. We notice that all three elements of x_t^q depend on all three elements of x_{t-1}^q and on the associating elements of the vector of disturbances ε_t^q. The state space formulation (3) has state vector $\alpha_t = x_t^q$ with T and R given by (7) and with system matrices T and R given by (6) and $Z = I_3$, $H = 0$, $Q = \sigma_\varepsilon^2$ and $\eta_t = \varepsilon_t^q$.

For the AR(2) process, matrix T is the same as in (7) but with $\phi_3 = 0$, such that its first column reduces to a zero vector and R is the same matrix as for the AR(3) process in (7).

AR(4) process

For AR(p) processes of order $p > 3$ in the state space formulation, the state vector in (3) is extended with more lags of x_t^q. In case of the AR(4) process, we have $\alpha_t = \left(x_{t-1,3}^q, x_{t,1}^q, x_{t,3}^q, x_{t,3}^q\right)'$ and we have

$$T = \begin{pmatrix} 0 & 0 & 0 & 1 \\ \phi_4 & \phi_3 & \phi_2 & \phi_1 \\ \phi_1\phi_4 & \phi_1\phi_3 + \phi_4 & \phi_1\phi_2 + \phi_3 & \phi_1^2 + \phi_2 \\ \phi_1^2\phi_4 + \phi_2\phi_4 & \phi_1^2\phi_3 + \phi_2\phi_3 + \phi_1\phi_4 & \phi_1^2\phi_2 + \phi_1\phi_3 + \phi_2^2 + \phi_4 & \phi_1^3 + 2\phi_1\phi_2 + \phi_3 \end{pmatrix}, \tag{8}$$

where the variance matrix of the process is defined as $\sigma_\varepsilon^2 RR'$ with

$$R = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \phi_1 & 1 & 0 \\ 0 & \phi_1^2 + \phi_2 & \phi_1 & 1 \end{pmatrix}. \tag{9}$$

For higher order AR(p) processes the same principle can be used. The state vector α_t has to be extended with the necessary lags of x_t^q and rows and columns with ones and zeros need to be added to the matrices T and R.

2.4 Computing Times

The AR(p) models can be specified as a linear Gaussian state space model (3) as we have illustrated in the previous section. Whether the monthly AR model is represented as the

monthly process x_τ^m, or as the stacked quarterly 3×1 vector x_t^q or as the yearly 12×1 vector x_t^y has no effect on the value of the loglikelihood function for given parameters $\phi_1, \ldots, \phi_p$ and σ_ε^2. The different model representations are observationally equivalent. In all cases, the Kalman filter can be used for likelihood evaluation. Hence the maximized loglikelihood value and the corresponding parameter estimates are the same for each case.

However, the representation has an effect on computing times. For example, when the data are treated as monthly observations, we have $n = 12,000$. When the data is stacked into quarterly 3×1 vectors we have $n = 4,000$ and with yearly 12×1 vectors we only have $n = 1,000$. Furthermore, with different frequencies we have different state vector lengths, and hence the transition matrix T and the variance matrix $\sigma_\varepsilon^2 RR'$ of the state vector have different dimensions for the different frequencies. The different dimensions will have a clear effect on the computations for the loglikelihood evaluation.

To illustrate this, we have evaluated the loglikelihood value 10,000 times for AR(p) models of different orders p and using the three different representations: monthly, quarterly and yearly. For each AR(p) process the calculations were performed on a time series with an AR(p) data generating process, consisting of 12,000 monthly observations. The loglikelihood value was calculated 10,000 times using the parameter values that maximized the loglikelihood function. We have verified that all likelihood evaluations resulted in the same value.

The computing times for the different combinations of AR(p) processes and frequencies are shown in Table 1. It is clear that for AR(1) and AR(2) processes, the structure with monthly single observations is most efficient. In these situations the smaller state vector outweighs the fact that the Kalman Filter has to go through 12,000 iterations instead of 4,000 and 1,000 for the quarterly and yearly structure, respectively. For AR(3) and higher orders, the quarterly structure is faster than the monthly. This can be explained by the fact that both structures have the same vector length for AR(3) and higher but the time dimension is three times smaller in the quarterly structure. For AR(10) and higher orders, the yearly structure is the fastest. For these orders the smaller time dimension starts to outweigh the larger state vector and for AR(12) and higher, all three structures have the same vector length, due to the number of lags that has to be included in the state vector in the monthly and quarterly structure. This example clearly illustrates that stacking observations into vectors with a lower frequency can lead to large gains in computational efficiency, especially when many lags of the observations are included in the model.

Table 1: Computing times

	Computing time (in seconds)			State vector length		
p	Monthly ($n = 12$k)	Quarterly ($n = 4$k)	Yearly ($n = 1$k)	Monthly ($n = 12$k)	Quarterly ($n = 4$k)	Yearly ($n = 1$k)
1	10	13	61	1	3	12
2	11	16	67	2	3	12
3	26	18	76	3	3	12
4	41	27	85	4	4	12
5	59	40	92	5	5	12
6	83	56	100	6	6	12
7	106	73	108	7	7	12
8	129	90	116	8	8	12
9	154	111	124	9	9	12
10	191	137	133	10	10	12
11	226	162	139	11	11	12
12	265	190	146	12	12	12

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

3 Mixed Frequency Dynamic Factor Models

In the previous section we have shown that in some situations computing times can be reduced by stacking the monthly data into observation vectors with a lower frequency. This is especially the case when many lags of the dependent variable are included in the model. Another situation in which stacking the data into vectors with a lower frequency is very convenient is when we simultaneously analyze variables which are observed at different frequencies. We then may stack the series observed at higher frequencies into vectors of the lowest frequency at which one of the variables is observed.

Suppose we have a monthly observed series x_τ^m, which is modeled by the AR(1) process (4) with $p = 1$, that is $x_\tau^m = \phi_x x_{\tau-1}^m + \varepsilon_\tau^m$ with the autoregressive parameter ϕ_x. We also have a quarterly observed series y_t, which is modeled by the AR(1) process as given by $y_t = \phi_y y_{t-1} + \xi_t$, with $\xi_t \sim NID\left(0, \sigma_\xi^2\right)$, where ϕ_y is the autoregressive coefficient for the quarterly lagged dependent variable y_{t-1} and ξ_t is the Gaussian disturbance that is possibly correlated with ε_t^q in the formulation of (5) for x_τ^m. We do not use the superscript q for y_t and ξ_t, because we assume throughout this section that they are specifically quarterly variables for which no monthly values become available.

When we stack the values of x_τ^m into quarterly 3×1 vectors x_t^q, as we have discussed in Section 2, then the two processes for y_t and x_t^q can be combined into a low frequency

multivariate process

$$\begin{pmatrix} y_{t+1} \\ x^q_{t+1,1} \\ x^q_{t+1,2} \\ x^q_{t+1,3} \end{pmatrix} = \begin{pmatrix} \phi_y & 0 & 0 & 0 \\ 0 & 0 & 0 & \phi_x \\ 0 & 0 & 0 & \phi_x^2 \\ 0 & 0 & 0 & \phi_x^3 \end{pmatrix} \begin{pmatrix} y_t \\ x^q_{t,1} \\ x^q_{t,2} \\ x^q_{t,3} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \phi_x & 1 & 0 \\ 0 & \phi_x^2 & \phi_x & 1 \end{pmatrix} \begin{pmatrix} \xi_t \\ \varepsilon^q_{t,1} \\ \varepsilon^q_{t,2} \\ \varepsilon^q_{t,3} \end{pmatrix}, \tag{10}$$

for $t = 1, \ldots, n$. We notice the difference between the autoregressive parameters ϕ_x and ϕ_y. The parameter ϕ_x measures the dependence of x_τ on its lagged value $x_{\tau-1}$ of one month earlier, whereas the parameter ϕ_y indicates the dependence of y_t on its lagged value y_{t-1} of one quarter earlier.

In the example above the two processes y_t and x^q_t are independent of each other, but the zeros in the first row and/or column of the matrix R can be replaced by non-zero values to include covariances between ξ_t and ε^q_t. These can be included if there is a relation between the contemporaneous values of y_t and x^q_t. Furthermore, the first row and column of the transition matrix T can be used to include lagged dependencies between the two series. The first row can be used when y_t is dependent on x^q_{t-1} and the first column if the values of x^q_t depend on the value of y_{t-1}. If other dynamic dependencies need to be introduced, we can extend formulations such as (10) straightforwardly.

3.1 Dynamic Factor Models

It is an increasingly common practice to simultaneously model y_t and x^m_τ by means of an unobserved common dynamic factor. The typical example is to have a large vector of x^m_τ variables that are potentially useful for the forecasting of y_t. For illustrative purposes, we initially assume that x^m_τ is a single variable. We then introduce a dynamic factor with a monthly frequency and we denote this factor by f^m_τ. The monthly variable x^m_τ has loading β_x on this common factor. This process can be described as

$$x^m_\tau = \beta_x f^m_\tau + \varepsilon^m_\tau, \qquad \varepsilon^m_\tau \sim NID\left(0, \sigma^2_\varepsilon\right), \tag{11}$$

where the common factor f^m_τ is specified by an AR(p) process, using the formulation

$$f^m_\tau = \phi_{f,1} f^m_{\tau-1} + \phi_{f,2} f^m_{\tau-2} + \ldots + \phi_{f,p} f^m_{\tau-p} + \eta^m_\tau, \qquad \eta^m_\tau \sim NID\left(0, \sigma^2_\eta\right), \tag{12}$$

The quarterly variable y_t is also loaded on the common factor f^m_τ, with loading β_y on all three values of f^m_τ within the quarter t. This can be described as

$$\begin{aligned} y_t &= \beta_y f^m_{(3t-2)} + \beta_y f^m_{(3t-1)} + \beta_y f^m_{(3t)} + \xi_t \\ &= \begin{pmatrix} \beta_y & \beta_y & \beta_y \end{pmatrix} f^q_t + \xi_t, \qquad\qquad \xi_t \sim NID\left(0, \sigma^2_\xi\right), \end{aligned} \tag{13}$$

where f^q_t is the 3×1 vector of stacked observations of the process f^m_τ using the stacking approach that was described for the series x^m_τ in (1).

Hence, the simultaneous process for y_t and x^m_τ can be described by a process in which

the values of x_τ^m and f_τ^m are stacked into the quarterly 3×1 vectors x_t^q and f_t^q

$$\begin{pmatrix} y_t \\ x_{t,1}^q \\ x_{t,2}^q \\ x_{t,3}^q \end{pmatrix} = \begin{pmatrix} \beta_y & \beta_y & \beta_y \\ \beta_x & 0 & 0 \\ 0 & \beta_x & 0 \\ 0 & 0 & \beta_x \end{pmatrix} \begin{pmatrix} f_{t,1}^q \\ f_{t,2}^q \\ f_{t,3}^q \end{pmatrix} + \begin{pmatrix} \xi_t \\ \varepsilon_{t,1}^m \\ \varepsilon_{t,2}^m \\ \varepsilon_{t,3}^m \end{pmatrix} \tag{14}$$

where the variance matrix of the disturbances is a 4×4 diagonal matrix with the first diagonal element equal to σ_ξ^2 and the other three diagonal elements equal to σ_ε^2. The vector autoregressive process for f_t^q is given by

$$f_{t+1}^q = T_f f_t^q + R_f \eta_t^q, \qquad \eta_t^q \sim NID\left(0, \sigma_\eta^2\right), \tag{15}$$

It is straightforward to generalize the model for y_t by having different loadings $\beta_{y,1}$, $\beta_{y,2}$ and $\beta_{y,3}$ for the three values of f_t^q within the quarter associated with t. Additional (monthly and quarterly) variables can also be included into the model and the model can be extended with specific factors for specific variables. The generality of the state space framework of Section 2.2 can be fully exploited.

We further can generalize the current specification by considering x_τ^m as a $k \times 1$ vector of variables. Mariano and Murasawa (2003) analyze such a dynamic factor model with one quarterly observed variable (in their case GDP growth) and four ($k = 4$) monthly observed economic indicators. In their analysis, they extend the Stock and Watson (1990) coincident index for business cycles, which uses four monthly indicators, with quarterly observed real GDP growth. Here we adopt this modeling framework but our analysis is based on the low frequency stacking approach as indicated above. In the empirical illustration of Section 5 we investigate different approaches for the simultaneous modeling of the five time series with a mix of quarterly and monthly frequencies.

First, in Section 3.2 we discuss the original solution by Mariano and Murasawa (2003) in treating mixed frequency data; they analyze all series at the highest frequency. In this approach, artificial missing values are introduced for the series that is observed at the lower frequency and interpolation techniques are used to describe the dynamics of the unobserved 'latent' monthly GDP growth. In Section 3.3 we explore a second approach where all series are modeled at the lowest frequency by aggregating the high frequency series to quarterly totals. In this approach, no artificial missing values are needed, but information about the high frequency series is lost and the model does not allow the econometrician to address high frequency dynamics. Finally, in Section 3.4 we discuss our model in which the high frequency series are stacked into vectors of observations at the lower frequency, as discussed in Section 2. In this approach, no artificial missing values are needed and also no information is lost about the high frequency series and their dynamic features.

3.2 Interpolation Approach

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

$$\begin{bmatrix} \cdot & \cdot & y_3 & \cdot & \cdot & y_6 & \cdot & \dots & y_{3n} \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & \dots & x_{3n} \end{bmatrix}, \tag{16}$$

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

$$\begin{pmatrix} \tilde{y}_\tau^m \\ x_\tau \end{pmatrix} = \begin{pmatrix} \beta_y g\left(f_\tau\right) \\ \beta_x f_\tau \end{pmatrix} + \begin{pmatrix} g(u_\tau) \\ v_\tau \end{pmatrix} + \begin{pmatrix} \xi_\tau \\ \varepsilon_\tau \end{pmatrix} \tag{17}$$

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

$$g(a_\tau) = \frac{1}{3}a_\tau + \frac{2}{3}a_{\tau-1} + a_{\tau-2} + \frac{2}{3}a_{\tau-3} + \frac{1}{3}a_{\tau-4} \tag{18}$$

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

The factors f_τ, u_τ and v_τ are modeled as AR processes

$$f_\tau \sim \mathrm{AR}(p_f), \qquad u_\tau \sim \mathrm{AR}(p_u), \qquad v_\tau \sim \mathrm{AR}(p_v), \tag{19}$$

where AR(p) refers to the process in (4) with order p that can be different, that is $p = p_f, p_u, p_v$ respectively. There are no interactions between the series of f_τ, u_τ and v_τ nor between the series of $v_\tau^{(i)}$ and $v_\tau^{(j)}$ for any $i \neq j$.

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

3.3 Aggregation Approach

An alternative approach, where the introduction of artificial missing values is not required, is the aggregation of the monthly series x_τ^m into quarterly totals $\bar{x}_t^q$ and the treatment of

all series as quarterly series. This model then only describes quarterly dynamics. The unobserved common factor must also become quarterly. To avoid cumbersome notation we drop the superscript q and we write the model in the form

$$\begin{pmatrix} y_t \\ \bar{x}_t \end{pmatrix} = \begin{bmatrix} \beta_y & 1 & 0 \\ \beta_x & 0 & I_k \end{bmatrix} \begin{pmatrix} f_t \\ u_t \\ \bar{v}_t \end{pmatrix} + \begin{pmatrix} \xi_t \\ \bar{\varepsilon}_t \end{pmatrix}, \tag{20}$$

for $t = 1, \ldots, n$, where $\bar{x}_t$ and β_x are both $k \times 1$ vectors. The common factor f_t and the idiosyncratic factors u_t and $\bar{v}_t$ can still be modeled by AR processes as in (19). However, we must take care when interpreting the values of the parameters of these processes, as they now describe the dynamics from quarter to quarter. We will hereafter refer to this approach as the mixed frequency aggregation (MFA) approach.

3.4 Stacking Approach

Here we introduce the stacking method of Section 2 for the mixed frequency dynamic factor model in the spirit of Mariano and Murasawa (2003). We consider for each quarter the observed values of y_t and x_t^q, where x_t^q consists of three stacked observations of the monthly vector variable x_τ^m, which we define as the $k \times 1$ vector $\left(x_\tau^{(1)}, \ldots, x_\tau^{(k)}\right)'$. In this approach, no artificial missing values are needed and no information is lost regarding the high frequency series and high frequency dynamics. When we drop the superscript q from the stacked vector x_t^q, we obtain the following model

$$\begin{pmatrix} y_t \\ x_{t,1}^{(1)} \\ x_{t,2}^{(1)} \\ x_{t,3}^{(1)} \\ \vdots \\ x_{t,1}^{(k)} \\ x_{t,2}^{(k)} \\ x_{t,3}^{(k)} \end{pmatrix} = \begin{bmatrix} \beta_y & \beta_y & \beta_y & 1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ \beta_{x^{(1)}} & 0 & 0 & 0 & 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & \beta_{x^{(1)}} & 0 & 0 & 0 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & \beta_{x^{(1)}} & 0 & 0 & 0 & 1 & \ldots & 0 & 0 & 0 \\ & \vdots & & & \vdots & & & \ddots & & \vdots & \\ \beta_{x^{(k)}} & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 0 & 0 \\ 0 & \beta_{x^{(k)}} & 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 1 & 0 \\ 0 & 0 & \beta_{x^{(k)}} & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} f_{t,1}^q \\ f_{t,2}^q \\ f_{t,3}^q \\ u_t \\ v_{t,1}^{(1)} \\ v_{t,2}^{(1)} \\ v_{t,3}^{(1)} \\ \vdots \\ v_{t,1}^{(k)} \\ v_{t,2}^{(k)} \\ v_{t,3}^{(k)} \end{pmatrix} + \begin{pmatrix} \xi_t \\ \varepsilon_{t,1}^{(1)} \\ \varepsilon_{t,2}^{(1)} \\ \varepsilon_{t,3}^{(1)} \\ \vdots \\ \varepsilon_{t,1}^{(k)} \\ \varepsilon_{t,2}^{(k)} \\ \varepsilon_{t,3}^{(k)} \end{pmatrix} \tag{21}$$

for $t = 1, \ldots, n$. Each vector series $x_t^{(i)}$ and $v_t^{(i)}$ have dimension 3×1 vectors; more formally we can write $x_t^{(i)} = \left(x_{t,1}^{(i)}, x_{t,2}^{(i)}, x_{t,3}^{(i)}\right)'$ and $v_t^{(i)} = \left(v_{t,1}^{(i)}, v_{t,2}^{(i)}, v_{t,3}^{(i)}\right)'$. The quarterly observed y_t has the same loadings β_y on all three elements of f_t^q, which is the 3×1 vector of stacked values of the monthly unobserved dynamic factor f_τ^m. Each element of $x_t^{(i)}$ depends on the corresponding element of the vector f_t^q with loading $\beta_{x^{(i)}}$, for $i = 1, \ldots, k$.

We can also opt for a quarterly unobserved common factor f_t so that we only have one

value for each quarter, for all three months. In this case, we do not use the superscript q and the model equation becomes

$$\begin{pmatrix} y_t \\ x_{t,1}^{(1)} \\ x_{t,2}^{(1)} \\ x_{t,3}^{(1)} \\ \vdots \\ x_{t,1}^{(k)} \\ x_{t,2}^{(k)} \\ x_{t,3}^{(k)} \end{pmatrix} = \begin{bmatrix} \beta_y & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \beta_{x^{(1)}} & 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ \beta_{x^{(1)}} & 0 & 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ \beta_{x^{(1)}} & 0 & 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & & & \vdots & & \ddots & & \vdots & \\ \beta_{x^{(k)}} & 0 & 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ \beta_{x^{(k)}} & 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ \beta_{x^{(k)}} & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} f_t \\ u_t \\ v_{t,1}^{(1)} \\ v_{t,2}^{(1)} \\ v_{t,3}^{(1)} \\ v_{t,1}^{(2)} \\ \vdots \\ v_{t,3}^{(k)} \end{pmatrix} + \begin{pmatrix} \xi_t \\ \varepsilon_{t,1}^{(1)} \\ \varepsilon_{t,2}^{(1)} \\ \varepsilon_{t,3}^{(1)} \\ \varepsilon_{t,1}^{(2)} \\ \vdots \\ \varepsilon_{t,3}^{(k)} \end{pmatrix} \tag{22}$$

where f_t is now a univariate quarterly dynamic factor. Here the quarterly observed variable y_t depends on the factor with loading β_y while each element of $x_t^{(i)}$ also depends on f_t and with the same loading $\beta_{x^{(i)}}$, for $i = 1, \dots, k$. We will hereafter refer to our solution as the Mixed Frequency Stacking (MFS) method. The former case of (21) has a monthly dynamic factor and is labelled as MFS-M while the latter case of (22) has a quarterly factor and is labelled as MFS-Q.

For both MFS-M and MFS-Q models, the dynamic factor f_t and the idiosyncratic factors u_t and v_t are modeled as AR processes as in (19). There are no interactions between the series of f_t, u_t and v_t nor between the series of $v_t^{(i)}$ and $v_t^{(j)}$ for any $i \neq j$.

3.5 Generalizations

The mixed frequency dynamic factor model with monthly observed variables x_τ^m that is stacked into the quarterly variable x_t^q and a quarterly observed variable y_t, is a specific example in which our stacking approach can be used to simultaneously model variables with mixed frequencies. We emphasize that the use of the stacking approach is not limited to this particular example. The approach can handle multiple variables with multiple different frequencies, as long as the observations of all variables can be stacked into vectors that have the lowest frequency. For example, if one or more yearly observed variables z_t^y would need to be added to the model, then all variables would have to be stacked into yearly vectors of observations. It is also possible to include variables with higher frequencies, such as daily or weekly. It only requires a somewhat more elaborate representation in state space form.

4 Weighted Maximum Likelihood Estimator

Next we introduce a weighted likelihood function for the mixed frequency dynamic factor model and we develop the asymptotic properties of the estimator that maximizes the weighted likelihood function. In this section we consider the basic dynamic factor model (11), (12) and (13) where both y_t and x_τ^m can be treated as vectors. The same

developments can be easily adapted for more general dynamic factor specifications. The loglikelihood function for the model can be given by

$$\mathcal{L}_T(\psi, f_1^m) := \log p(y, x; \psi) = \log p(y|x; \psi) + \log p(x; \psi), \tag{23}$$

where p is the Gaussian density function, stacked vectors y and x collect all T observations available for the variables y_t and x_τ^m in the sample, respectively, the initial value of the monthly unobserved dynamic factor f_1^m is treated as a fixed value, parameter vector ψ collects all unknown coefficients in the model including the factor loadings β_x and β_y, autoregressive parameters for the AR process, and the variances of the disturbances. The loglikelihood expression shows that the joint density can alternatively be expressed by a conditional density multiplied by a marginal density. For our purpose this expression 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_τ^m are typically instrumental to improve the nowcasts and forecasts of y_t.

Under the assumption that y and x are jointly generated by a Gaussian dynamic factor model such as (11), (12) and (13), we can apply the Kalman filter to evaluate the log-likelihood function via the prediction error decomposition. Koopman and Durbin (2000) discuss an alternative filtering method in which each element of the observation vector $(y', x')'$ is brought one at a time into the updating step of the Kalman filter. In effect, the vector series is converted into a univariate time series where multiple observations are available for the same time index. In our application, it means that for each available observation, whether it is for a monthly or a quarterly variable, an update takes place. It results in a truly online and high frequency filtering method even though the model is formulated in terms of a low frequency time index.

The Kalman filter is applied to the low frequency model with the (quarterly) time index t and for the $r_t \times 1$ observation vector $z_t := (y_t, x_t')'$ where x_t is the stacked vector of monthly variables x_τ^m within the quarter associated with time t, for $t = 1, \ldots, n$. Hence the total number of observations is $T = \sum_{t=1}^n r_t$. The loglikelihood function is then given by

$$\mathcal{L}_T(\psi, f_1^m) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^n \ell_t(z_t|\tilde{f}_t(\tilde{f}_1); \psi), \tag{24}$$

where ℓ_t is the predictive Gaussian logdensity contribution at time t and $\tilde{f}_t(\tilde{f}_1)$ is the predictor of the unobserved factor f_t^m conditional on past observations $z_1, \ldots, z_{t-1}$ and its initial value $\tilde{f}_1 = f_1^m$. The Kalman filter evaluates recursively $\tilde{f}_t(\tilde{f}_1)$ from which ℓ_t can be evaluated. In spirit of the likelihood decomposition in (23), likelihood evaluation via the method of Koopman and Durbin (2000) is partly based on

$$\mathcal{L}_T(\psi, f_1^m) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^n \ell_t(y_t|\tilde{f}_t(\tilde{f}_1); \psi) + -\frac{1}{2}\sum_{t=1}^n \ell_t(x_t|y_t, \tilde{f}_t(\tilde{f}_1); \psi), \tag{25}$$

which facilitates a separate treatment of y_t and x_t.

Maximum likelihood estimation for the parameter vector ψ is based on applying a numerical quasi-Newton optimization method for the maximization of $\mathcal{L}_T(\psi, f_1^m)$, with respect to ψ. The maximization is an iterative process. After 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 the mixed frequency dynamic factor model, the treatment of the observations in z_t for the construction of the likelihood function is implied by the dynamic factor model. However, it is very likely that the dynamic factor model is misspecified as a model representation of the true data generation process for the variables represented in z_t. When our primary aim is to analyze y_t in particular, we may be less concerned with the misspecification of x_t, to some extent. To reflect the higher importance of y_t in comparison to x_t in the likelihood construction for the misspecified dynamic factor model, we propose to give different weights to the likelihood contributions of y_t and x_t explicitly. Hence we propose the weighted loglikelihood function

$$\mathcal{L}_T(\psi, w, f_1^m) = -\frac{T}{2}\log(2\pi) - \frac{W}{2}\sum_{t=1}^{n}\ell_t(y_t|\tilde{f}_t(\tilde{f}_1);\psi) + -\frac{1}{2}\sum_{t=1}^{n}\ell_t(x_t|y_t,\tilde{f}_t(\tilde{f}_1);\psi), \quad (26)$$

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 (26) is referred to as the weighted maximum likelihood (WML) estimator.

4.1 Asymptotic Properties of the WML Estimator

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

Proposition 1 below states well known conditions for the strict stationarity and ergodicity (SE) of the true processes $\{f_\tau^m\}_{\tau\in\mathbb{Z}}$, $\{x_\tau^m\}_{\tau\in\mathbb{Z}}$ and $\{y_t\}_{t\in\mathbb{Z}}$ generated by the linear Gaussian system in (11), (12) and (13) initialized in the infinite past.

Proposition 1. *Let $\{x_\tau^m\}_{\tau\in\mathbb{Z}}$ and $\{y_t\}_{t\in\mathbb{Z}}$ be generated according to (11), (12) and (13)*

with

(i) $\|T_f\| < 1$, $0 < \|R_f\| < \infty$ *and* $0 < \|\Sigma_\eta^2\| < \infty$ *in (11);*

(ii) $|\beta_x| < \infty$ *and* $0 < \sigma_\epsilon^2 < \infty$ *in (12);*

(iii) $|\beta_y| < \infty$ *and* $0 < \sigma_\zeta^2 < \infty$ *in (13).*

Then $\{x_\tau^m\}_{\tau\in\mathbb{Z}}$ *and* $\{y_t\}_{t\in\mathbb{Z}}$ *are SE sequences with bounded moments of any order; i.e.* $\mathbb{E}|x_\tau^m|^r < \infty$ *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* $\hat{\psi}_T(w, \tilde{f}_1^m) : \Omega \to \Psi$ *satisfying*

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

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

Theorem 2 establishes the strong consistency of the WML estimator of the true parameter vector $\psi_0 \in \Psi$ for any choice of weight $w \in (0,1]$ for the likelihood. This result is obtained under the assumption that the mixed frequencies common factor model is well-specified and for any filter that identifies the parameter vector $\psi_0 \in \Psi$ and is asymptotically SE with bounded moments of second order. The identification conditions and exponential almost sure (e.a.s.) convergence of different filters to an SE process with bounded second moment is well known and easy to establish in this linear Gaussian setting. For this reason, we do not repeat them here; see e.g. Mehra (1970) for such results on the classical Kalman filter, Bougerol (1993) for extensions, and Blasques, Koopman, and Lucas (2014) for identification, convergence results and bounded moments on a wide range of observation-driven filters. Theorem 2 thus assumes that ψ_0 maximizes the likelihood and assumes the convergence of the filtered sequence $\{\tilde{f}_\tau^m(\tilde{f}_1^m)\}_{\tau\in\mathbb{N}}$ initialized at $\tilde{f}_1^m$ to a unique limit SE sequence $\{\tilde{f}_\tau^m\}_{\tau\in\mathbb{Z}}$ with bounded second moment. Notice that we just require identification in the usual ML setting $w = 1$; i.e. identification w.r.t. the unweighted likelihood function $\mathcal{L}_T(\psi, 1)$. As shown in the proof, 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* $\{x_\tau^m\}$ *and* $\{y_t\}$ *be generated by the mixed frequencies common factor model defined in (11), (12) and (13) under some* $\psi_0 \in \Psi$, *and suppose that the conditions of Propositions 1 and Theorem 1 hold. Suppose furthermore that*

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

and there exists a unique SE sequence such that

$$\|\tilde{f}_\tau^m(\tilde{f}_1^m) - \tilde{f}_\tau^m\| \overset{e.a.s.}{\to} 0 \ \forall\ \tilde{f}_1^m \quad as \quad \tau \to \infty \quad with \quad \mathbb{E}|\tilde{f}_\tau^m|^2 < \infty.$$

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

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

for any choice of weight $w \in (0, 1]$ *and any initialization* $\tilde{f}_1^m$.

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

It is well known that the classical ML estimator converges to a limit pseudo-true parameter that minimizes the KL divergence between the true joint probability measure of the data and the measure implied by the model. Theorem 3 characterizes the limit pseudo-true parameter $\psi_0^*(w)$ as the minimizer of a transformed KL divergence for every given $w \in (0, 1]$. Just like 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_τ, and bounded from below (for $w = 0$) by the conditional density of y_t given x_τ^m. 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_τ^m, 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 $\mathrm{KL}(a, b)$ denote the KL divergence between a and b.

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

$$\mathcal{L}_\infty(\psi_0^*(w), w) > \mathcal{L}_\infty(\psi, w) \ \forall \ \psi \neq \psi_0^*(w)$$

and there exists a unique SE sequence such that

$$\|\tilde{f}_\tau(\tilde{f}_1^m) - \tilde{f}_\tau\| \xrightarrow{e.a.s.} 0 \ \forall \ \tilde{f}_1^m \quad as \quad \tau \to \infty \qquad with \qquad \mathbb{E}|\tilde{f}_\tau|^2 < \infty.$$

Then

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

for any initialization $\tilde{f}_1^m$ *and any weight* $w \in (0, 1]$. *Furthermore, the pseudo-true param-*

eter $\psi_0^*(w)$ *minimizes a transformed KL divergence*

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

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

$$\mathrm{TKL}_1(q(\cdot;\psi),p) = \mathrm{KL}(q(\cdot;\psi),p) \quad , \quad \mathrm{TKL}_0(q(\cdot;\psi),p) = \mathrm{KL}(q_1(\cdot;\psi),p_1) \ ,$$

$$\mathrm{KL}(q_1(\cdot;\psi),p_1) \;\le\; \mathrm{TKL}(q(\cdot;\psi),p) \;\le\; \mathrm{KL}(q(\cdot;\psi),p) \ ,$$

$$\textit{and} \quad \mathrm{TKL}(q(\cdot;\psi),p) = 0 \quad \textit{if and only if} \quad \mathrm{KL}(q_1(\cdot;\psi),p_1) = 0.$$

Theorem 4 establishes the asymptotic normality of the WML estimator under the assumption that the mixed frequencies dynamic factor model is well specified. Below we let $\mathcal{J}(\psi_0,w) := \mathbb{E}\ell_t'(\psi_0,w)\ell_t'(\psi_0,w)^\top$ denote the expected outer product of gradients and $\mathcal{I}(\psi_0,w) := \mathbb{E}\ell_t''(\psi_0,w)$ be the Fisher information matrix. The asymptotic normality proof is written for filters whose derivative processes are asymptotically SE and have bounded moments; see Blasques et at. (2014) for a wide range of observation-driven filters satisfying such conditions. Below, $\{\tilde{df}_\tau^m(\tilde{df}_1^m)\}$ and $\{d\tilde{df}_\tau^m(d\tilde{df}_1^m)\}$ denote the first and second derivatives of the filter w.r.t. the parameter vector ψ, initialized at $\tilde{df}_1^m$ and $d\tilde{df}_1^m$ respectively. Their SE limits are denoted $\{\tilde{df}_\tau^m\}$ and $\{d\tilde{df}_\tau^m\}$. 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* $\{\tilde{df}_\tau^m\}$ *such that*

$$\|\tilde{df}_\tau^m(\tilde{df}_1^m) - \tilde{df}_\tau^m\| \overset{e.a.s.}{\to} 0 \;\forall\; \tilde{df}_1^m \quad as \quad \tau \to \infty \quad with \quad \mathbb{E}|\tilde{df}_\tau^m|^4 < \infty$$

and a unique SE sequence $\{d\tilde{df}_\tau\}$ *such that*

$$\|d\tilde{df}_\tau^m(d\tilde{df}_1^m) - d\tilde{df}_\tau^m\| \overset{e.a.s.}{\to} 0 \;\forall\; d\tilde{df}_1^m \quad as \quad \tau \to \infty \quad with \quad \mathbb{E}|d\tilde{df}_\tau^m|^2 < \infty.$$

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

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

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

Theorem 5. (Asymptotic Normality) *Let the conditions of Theorem 3 hold and* $\psi_0^*(w)$ *be a point in the interior of* Ψ. *Suppose further that* $\{x_\tau^m\}$ *and* $\{y_t\}$ *are SE and satisfy*

$\mathbb{E}|x_\tau^m|^4 < \infty$ *and* $\mathbb{E}|y_t|^4 < \infty$ *and there exists a unique SE sequence* $\{\tilde{df}_\tau^m\}$ *such that*

$$\|\tilde{df}_\tau^m(\tilde{df}_1^m) - \tilde{df}_\tau^m\| \overset{e.a.s.}{\to} 0 \;\forall\; \tilde{df}_1^m \quad as \quad \tau \to \infty \qquad with \qquad \mathbb{E}|\tilde{df}_\tau^m|^4 < \infty$$

and a unique SE sequence $\{d\tilde{df}_\tau\}$ *such that*

$$\|d\tilde{df}_\tau^m(d\tilde{df}_1^m) - d\tilde{df}_\tau^m\| \overset{e.a.s.}{\to} 0 \;\forall\; d\tilde{df}_1^m \quad as \quad \tau \to \infty \qquad with \qquad \mathbb{E}|d\tilde{df}_\tau^m|^2 < \infty.$$

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

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

4.2 Small Sample Properties of WML: Monte Carlo Study

Next we investigate the finite sample effects of different choices for the value of W in (26) on the in-sample fit in different scenarios using Monte Carlo simulations. We generate data for y_t and x_τ^m for different number of variables, $k = 2$, $k = 5$ and $k = 10$. The length of the time series is set to $n = 100$ for all scenarios. We consider two different data generating processes (DGPs) for the vector of observations $z_t = (y_t, x_t')'$ in the simulations, where x_t is the stacked vector of monthly variables x_τ^m.

The first DGP for z_t is a multivariate common factor model with one common factor f_τ and one idiosyncratic factor $u_\tau^{(i)}$ for each variable. The model is given by

$$z_t = \beta_z f_t + u_t + \varepsilon_t, \qquad \varepsilon_\tau \sim NID\left(0, \sigma_\varepsilon^2\right) \tag{27}$$

All factors, both common and idiosyncratic, are driven by AR processes of order 1 as in (19) and the autoregressive parameters are all set at 0.80. The factor loadings of variable $x^{(i)}$ on the common factor are set equal to $\frac{1}{i}$, for $i = 1, \ldots, k$. Furthermore, we choose $\sigma_\varepsilon = 0.50$ and $\sigma_\eta = 0.25$.

The second DGP for z_t is the multivariate VAR(1) process

$$z_t = \Phi z_{t-1} + \varepsilon_t, \qquad \varepsilon_\tau \sim NID\left(0, \sigma_\varepsilon^2\right) \tag{28}$$

where all diagonal elements of the $k \times k$ matrix Φ are set equal to 0.80 and the off-diagonal elements are randomly generated values between -0.5 and 0.5 but we only allow for stationary VAR processes. As in the first DGP, we choose $\sigma_\varepsilon = 0.50$. In our simulation study, we consider three different scenarios in particular.

Scenario 1: Underspecification In the first scenario, we adopt the common dynamic factor model (27) as the DGP but for the analysis we consider the model

$$z_t = \beta_z f_t + \varepsilon_t, \qquad \varepsilon_\tau \sim NID\left(0, \sigma_\varepsilon^2\right). \tag{29}$$

This model does not include idiosyncratic factors and hence it is underspecified.

We expect that in this setting we can improve the in-sample accuracy for y_t when we take the value of W larger than unity. For different dimensions, $k = 2$, $k = 5$ and $k = 10$, we generate 1,000 datasets from the first DPG and we estimate the parameters from the model (29) with values of W ranging from 1 to 25.

Scenario 2: Misspecification In the second scenario, we adopt the VAR(1) model (28) as DGP. Similarly as in the first scenario, we generate 1,000 datasets from the VAR(1) with $k = 2$, $k = 5$ and $k = 10$, and we estimate the parameters of the dynamic factor model (29) with values of W ranging from 1 to 25. Since the DGP model is different from the model that we consider for te analysis, the model is misspecified. Hence we expect that increasing the value of W can be beneficial for the in-sample fit of y_t. Theorem 3 has shown that such large improvements are explained by the fact that we can use the weight W to control the limit pseudo-true parameter to which the weighted MLE will converge. In particular, the larger the W, the more the WML estimator will focus on minimizing the KL divergence between the true conditional density of y_t and the conditional density implied by the model.

Scenario 3: Correct Specification In the third and final scenario, we adopt the dynamic factor (27) as DGP and we consider the same model in the analysis. We generate 1,000 datasets for $k = 2$, $k = 5$ and $k = 10$, and we estimate the parameters for values of W ranging from 1 to 25. This model is correctly specified and we expect that increasing the value of W will not improve the in-sample fit for y_t much in comparison to the previous two scenarios. In effect, in Theorem 2 we have shown that asymptotically the different values of W must yield the same results since the weighted MLE is consistent to the true parameter for any W. Any improvements in the correct specification setting are thus only finite-sample improvements.

In the left panel of Table 2 we present the mean squared error (MSE) averages for the variable of interest y_t. Each column is scaled to the value at $W = 1$, the maximum likelihood estimate. From Table 2 we find that increasing the value of W leads to a better in-sample fit for y_t for all three dimensions k. However it is not necessary to choose very large values for W. The improvements in MSE appear to converge to some upper limit for increasing values of W. To illustrate we also report the results for $W = 1,000$ from which we learn that there is not much difference with $W = 25$. Furthermore, we observe that more gains can be made when more variables are included in the model.

In the first panel of Figure 1 we present the MSE of y_t and the average MSEs for the variables of x_t for $W = 1, \ldots, 25$. As we have learned from Table 2, increasing the value of W improves the in-sample fit for y_t while the actual improvements diminish when W increases. It comes at the cost of the accuracy for the variables in x_t. The right panel of each figure presents the average MSE for x_t and it is clear that increasing W leads to a worse in-sample fit for these variables. We also find that the marginal effect becomes smaller for larger values of W. Since the MSE of y_t is decreasing as a function of W,

there is no clear optimal value for W. The choice of W should depend on how important the different variables are considered to be.

For scenario 2 we present the improvements in average MSE for y_t in the middle panel of Table 2. As in the first scenario, we observe that increasing the value of W leads to a better in-sample fit for y_t for all the three values of k, and again the improvement in MSE seems to converge to some upper limit for increasing values of W. This is consistent with the findings of Theorem 3. We can also conclude that the gains for $k = 5$ and $k = 10$ are much larger than in scenario 1. However, more gains can be made in the model with $k = 5$ than in the model with $k = 10$.

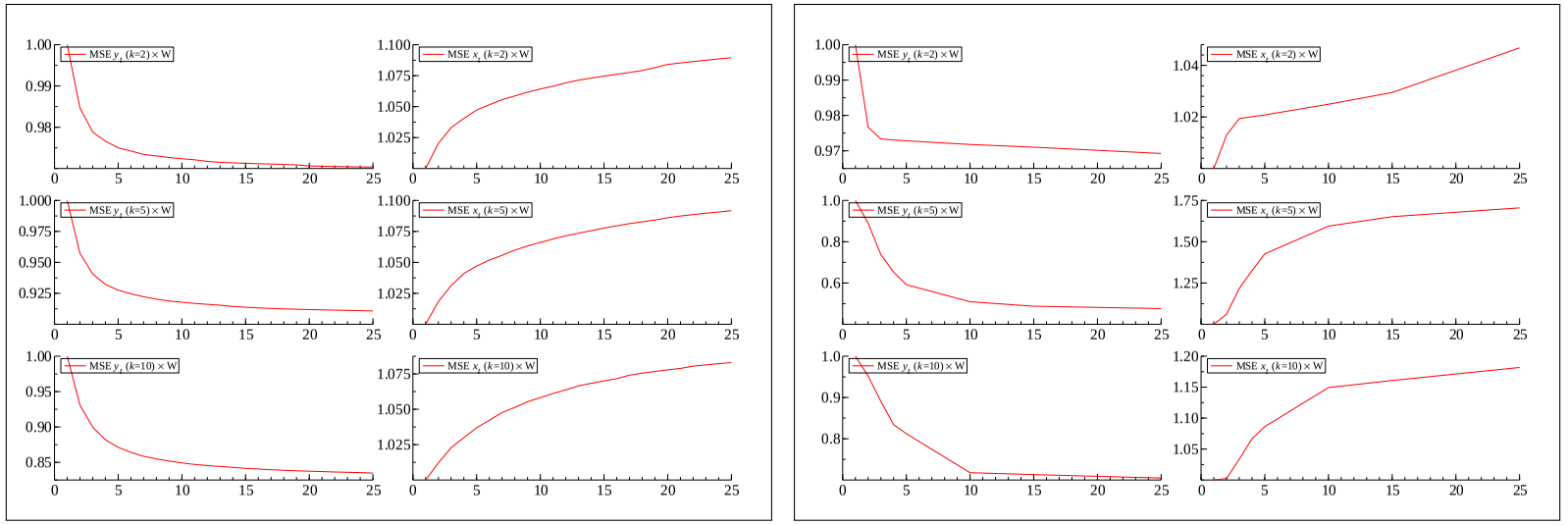
Table 2: Average MSE of target variable y_t for different values of W

	Scenario 1			Scenario 2			Scenario 3		
W	$k = 2$	$k = 5$	$k = 10$	$k = 2$	$k = 5$	$k = 10$	$k = 2$	$k = 5$	$k = 10$
1	1.000	1.000	1.000	1.000	1.000	1.000	1.0000	1.0000	1.0000
2	0.983	0.962	0.931	0.977	0.890	0.952	0.9996	0.9994	0.9999
3	0.974	0.947	0.889	0.973	0.737	0.891	0.9994	0.9994	0.9998
4	0.972	0.942	0.874	0.973	0.653	0.834	0.9993	0.9993	0.9997
5	0.970	0.938	0.865	0.973	0.592	0.812	0.9992	0.9993	0.9993
10	0.968	0.928	0.844	0.972	0.509	0.718	0.9990	0.9990	0.9989
15	0.967	0.925	0.837	0.971	0.488	0.713	0.9989	0.9989	0.9986
25	0.966	0.920	0.831	0.969	0.476	0.705	0.9988	0.9987	0.9984
1000	0.965	0.914	0.809	0.965	0.442	0.685	0.9986	0.9976	0.9958

This table presents the average MSE for the target variable y_t over 1,000 simulation runs. In scenario 1 an underspecified DFM (without idiosyncratic factors) is estimated on a DFM with idiosyncratic factors. In scenario 2 a DFM (without idiosyncratic factors) is estimated on a dataset with a VAR(1) process as DGP and in scenario 3 a DFM model with idiosyncratic factors is estimated on a dataset with the same model as the DGP. For each column all numbers are scaled to the value at $W = 1$.

Finally, for scenario 3 we present the results in the right panel of Table 2. The improvements in in-sample fit are negligible when the value of W increases. This is consistent with Theorem 2, as any improvements can only be attributed to small-sample variations.

Figure 1: MSE for Scenario 1 (Underspecification) and Scenario 2 (Misspecification)



This figure presents the in-sample MSE of the one-step ahead predictions for different values of W in the weighted loglikelihood function (26) for $k = 2$ (upper graphs), $k = 5$ (middle graphs) and $k = 10$ (lower graphs) variables. The numbers on the horizontal axis represent the value of W and all numbers are relative to the values at $W = 1$. In the first panel we use a misspecified model. The DGP is a VAR(1) model and the estimated model is a dynamic factor model with one common factor. In the second panel we use an underspecified model. The DGP is a common factor model with idiosyncratic factors and the estimated model is a model with only a common factor. The MSEs of the target variable y_t are presented in the graphs on the left within each panel. The graphs on the right present the average MSEs of the other variables x_t.

5 Empirical Results

In our empirical study we consider four different methods for treating mixed frequency dynamic factor models (DFMs): the MFI method of Section 3.2, the MFA method of Section 3.3 and the MFS-M and MFS-Q methods of Section 3.4. We first present the estimation results for the four models based on the original dataset that was used by Mariano and Murasawa (2003). We will evaluate the forecasting and nowcasting accuracy for the various methods and compare the results. These different approaches to handle mixed-frequencies with a dynamic factor model, are all examples of full system approaches. We will also compare the forecasting and nowcasting results with two partial model approaches: Bridge models of Trehan (1989) and the MIDAS regression of Ghysels, Santa-Clara, and Valkanov (2006). We refer for some details of these two approaches to the Technical Appendix. Furthermore we assess the improvements in forecasting and nowcasting accuracy when we adopt the method of weighted maximum likelihood that is presented in Section 4.

5.1 Data

The original Mariano and Murasawa (2003) data set consist of quarterly observed US real GDP and four monthly observed indicators; see Table 3 for their descriptions. The sample period is January 1959 until December 2000. The first difference of natural log for each series is taken and multiplied by 100, which is approximately equal to the quarterly or monthly growth rate. The number of parameters in the model is reduced by demeaning all series, so that no intercept coefficients are required. Furthermore, the loading of the dynamic factor on quarterly GDP is fixed at a value of 1 for identification purposes. In the forecasting and nowcasting studies, we use an extended version of this dataset, with the sample period from January 1960 until December 2009.

Table 3: Data definitions

Indicator	Description
	Quarterly
GDP	Real GDP (billions of chained 1996 $, SA, AR)
	Monthly
EMP	Employees on non-agricultural payrolls (thousands, SA)
INC	Personal income less transfer payments (billions of chained 1996 $, SA, AR)
IIP	Index of industrial production (1992 = 100, SA)
SLS	Manufacturing and trade sales (millions of chained $, SA)

This table presents the definitions of all the quarterly and monthly variables that are used in the models in this section and the abbreviations that are used for these variables. SA means 'seasonally adjusted' and AR means 'annual rate'.

5.2 Parameter Estimates by Maximum Likelihood

We have obtained our estimation results by following Mariano and Murasawa (2003) closely. We have set the variances of the disturbance terms in the observation equations of all models equal to zero and we also adopt the orders that they use for the AR(p) processes by which the factors are modeled. That is, the common factor f_t is modeled by an AR(1) process and all idiosyncratic factors u_t and v_t are modeled by AR(2) processes for all models.

The parameter estimates for MFI model are presented in the first panel of Table 4. These estimates are very similar to the estimates obtained by Mariano and Murasawa (2003) for the same dataset. The MFS-M model targets the same modeling framework as assumed for the MFI model. However, the MFI model describes the dynamic information for Δ ln GDP from a high frequency monthly process, while the MFS-M model adopts a low frequency quarterly dynamic process. Overall, we obtain similar estimation results for both methods. The only notable difference is in the estimated parameters $\phi_{u,1}$ and $\phi_{u,2}$ for Δ ln GDP. The third panel presents the parameter estimates for the MFS-Q model, which are also very similar to the estimates of the MFI and MFS-M methods. However, we notice that all elements of β_x are now estimated at about $\frac{1}{3}$ of their value in the MFI and MFS-M models: in the MFS-Q model this is the loading on the quarterly value of f_t, whereas in the MFI and MFS-M models it is the loading on the monthly value of f_t. Finally, the fourth panel reports the parameter estimates for the MFA model. Given that the number of observations for the quarterly model has reduced, the variances of the idiosyncratic factors have increased for all four indicators. The estimates of the other variables are quite similar to those of the MFS-M model, with the exception of the AR process for Δ ln EMP ($x^{(1)}$). This process has become rather persistent with an estimated $\phi_{1,1}$ of 0.69 when Δ ln EMP is modeled by a quarterly frequency. When it is modeled with a monthly frequency, the estimate of $\phi_{1,1}$ has been close to zero, while the $\phi_{1,2}$ estimates have been at 0.45 and 0.43 in the MFI and MFS-M models, respectively.

5.3 Nowcasting and Forecasting Comparisons

Next we compare the forecasting results of our three approaches at handling mixed-frequencies in the dynamic factor model. We also compare these results to the forecasting results that are obtained with bridge models and MIDAS regressions.

We compare the models using the same set of variables as in Mariano and Murasawa (2003), but with the data extended until December 2009. Table 5 presents the MSE of the predictions for the quarterly observed y_t in the different models with forecasting horizons $h = 0, 1, 2, 3, 6$ months ahead. These predictions are made using a rolling window of 25 years of data to estimate the parameters. Forecasting horizons of $h = 0, 1, 2$ are usually referred to as nowcasting. When $h = 0$, all values of x_t are known until the quarter that is to be forecasted. When $h = 1$, the values of x_t are known until the first two months of the quarter. When $h = 2$, only the first month of the quarter that we

Table 4: Parameter estimates

MFI Model

Parameter	Δ ln GDP	Δ ln EMP	Δ ln INC	Δ ln IIP	Δ ln SLS
β	1.00	0.49	0.81	2.14	1.74
		(0.04)	(0.06)	(0.13)	(0.11)
ϕ_F			0.56		
			(0.05)		
σ_F^2			0.08		
			(0.01)		
$\phi_{u,1}$	-0.04	0.10	-0.05	-0.05	-0.41
	(0.08)	(0.04)	(0.04)	(0.07)	(0.05)
$\phi_{u,2}$	-0.83	0.45	0.03	-0.06	-0.20
	(0.07)	(0.05)	(0.05)	(0.06)	(0.05)
$\sigma_{u,2}^2$	0.19	0.02	0.09	0.25	0.61
	(0.04)	(0.00)	(0.01)	(0.02)	(0.04)

MFS-M Model

Parameter	Δ ln GDP	Δ ln EMP	Δ ln INC	Δ ln IIP	Δ ln SLS
β	1.00	0.57	0.90	2.30	1.83
		(0.04)	(0.06)	(0.13)	(0.12)
ϕ_F			0.59		
			(0.04)		
σ_F^2			0.06		
			(0.01)		
$\phi_{u,1}$	-0.40	0.07	-0.08	-0.01	-0.38
	(0.09)	(0.05)	(0.05)	(0.05)	(0.05)
$\phi_{u,2}$	-0.21	0.43	0.01	-0.05	-0.17
	(0.16)	(0.06)	(0.07)	(0.07)	(0.07)
$\sigma_{u,2}^2$	0.27	0.02	0.09	0.27	0.64
	(0.04)	(0.00)	(0.01)	(0.03)	(0.05)

MFS-Q Model

Parameter	Δ ln GDP	Δ ln EMP	Δ ln INC	Δ ln IIP	Δ ln SLS
β	1.00	0.25	0.33	0.72	0.60
		(0.02)	(0.02)	(0.04)	(0.04)
ϕ_F			0.69		
			(0.06)		
σ_F^2			0.25		
			(0.04)		
$\phi_{u,1}$	-0.30	0.11	0.10	-0.10	-0.37
	(0.09)	(0.05)	(0.04)	(0.05)	(0.04)
$\phi_{u,2}$	-0.13	0.24	-0.06	-0.11	-0.20
	(0.13)	(0.07)	(0.05)	(0.06)	(0.06)
$\sigma_{u,2}^2$	0.24	0.03	0.10	0.34	0.74
	(0.03)	(0.00)	(0.01)	(0.02)	(0.05)

MFA Model

Parameter	Δ ln GDP	Δ ln EMP	Δ ln INC	Δ ln IIP	Δ ln SLS
β	1.00	0.67	0.95	2.18	1.77
		(0.06)	(0.08)	(0.12)	(0.11)
ϕ_F			0.68		
			(0.06)		
σ_F^2			0.26		
			(0.04)		
$\phi_{u,1}$	-0.27	0.69	-0.05	-0.14	-0.22
	(0.09)	(0.11)	(0.08)	(0.11)	(0.08)
$\phi_{u,2}$	-0.11	0.09	-0.03	-0.05	-0.19
	(0.12)	(0.11)	(0.10)	(0.14)	(0.10)
$\sigma_{u,2}^2$	0.25	0.06	0.40	0.56	1.11
	(0.03)	(0.01)	(0.05)	(0.10)	(0.13)

This table presents all parameter estimates for the the four different dynamic factor models that are evaluated in this section. The four dynamic factor models are Mixed frequency Interpolation (MFI), Mixed frequency Stacking with Monthly common factor (MFS-M), Mixed frequency Stacking with Quarterly common factor (MFS-Q) and Mixed frequency Aggregation (MFA). All models are estimated on a dataset with a sample period from January 1960 until December 2000 and using the data definitions as described in Table 3. Numbers in parentheses are asymptotic standard errors.

want to forecast is observed. When $h = 3$, we are forecasting one quarter ahead, no observations are available for the quarter that we forecast. All values until the previous quarter are observed. Similarly, when $h = 6$ we are forecasting two quarters ahead without observations being available from the quarter to forecast and its previous quarter. In the MFA model, there are no monthly dynamics, so it is only possible to forecast at $h = 0, 3, 6$.

From Table 3 we learn that the MFS-M model provides the more accurate Δ ln GDP predictions for all forecasting horizons. The accuracy is higher than the original MFI model. The MFA model is only more accurate for $h = 3$ and $h = 6$ while the MFS-Q model is most accurate for $h = 6$ while it is less accurate than both the MFI and MFS-M models for all other horizons.

When we further take into account the benchmark Bridge and MIDAS models, we find that the predictions of the MIDAS model are most accurate for nowcasting at $h = 0$ and $h = 1$. For longer forecasting horizons, MIDAS forecasts become less accurate.

We consult the Diebold and Mariano (1995) test to verify whether forecast accuracy differences are significant at the 5% level. In our study, the only significant differences between any of the models for the nowcasting horizons $h = 0, 1, 2$ is found for (i) $h = 2$, where the MFS-M model is more accurate than the MFS-Q model; (ii) $h = 3$, where the MFA model is significantly more accurate than the MFI and MIDAS models; (iii) $h = 6$, where the MFS-Q model is significantly more accurate than all other models and the MFA model is significantly more accurate than the rest of the models, except for the MFS-M model. Given the large similarity in the structure of these models and their parameter estimates, it is perhaps only surprising that we find any difference at all. Finally, we have found that the MIDAS model performs significantly worse than all of the other models for $h = 6$.

Table 5: Forecast comparisons for US GDP growth

	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 6$
MFI	0.1779	0.1918	0.2340	0.3156	0.4023
MFS-M	0.1666	0.1730	0.2108	0.2935	0.3986
MFS-Q	0.1765	0.1909	0.2411	0.2989	0.3701
MFA	0.1693			0.2809	0.3754
BM	0.1833	0.2056	0.2455	0.3046	0.4180
MIDAS	0.1597	0.1658	0.2464	0.3635	0.4873

This table presents the forecast MSEs for the quarterly observed US GDP growth from January 2000 until December 2009 for four dynamic factor models and two benchmark models at forecasting horizons $h = 0, 1, 2, 3, 6$. The four dynamic factor models are Mixed frequency Interpolation (MFI), Mixed frequency Stacking with Monthly common factor (MFS-M), Mixed frequency Stacking with Quarterly common factor (MFS-Q) and Mixed frequency Aggregation (MFA). Benchmark models are Bridge models (BM) and MIDAS regressions. All results are based on parameter estimates obtained from the same dataset with sample period from January 1960 until December 2009 and using the data definitions as described in Table 3. For each forecasting horizon the MSEs of the two most accurate models are highlighted, where the darkest shade is used for the most accurate model.

5.4 Parameter Estimates by Weighted Maximum Likelihood

Dynamic factor models are often considered in the simultaneous modeling of a group of variables, but all variables in the model are of equal importance to the likelihood function. However, it can be desirable to have somewhat more focus on fitting a particular variable of interest. In our case, this may be Δ ln GDP. In Section 4 we have proposed a method to implement this idea by giving extra weight to the most important variables in the log-likelihood function. In our model we will give extra weight to GDP growth, the variable y_t.

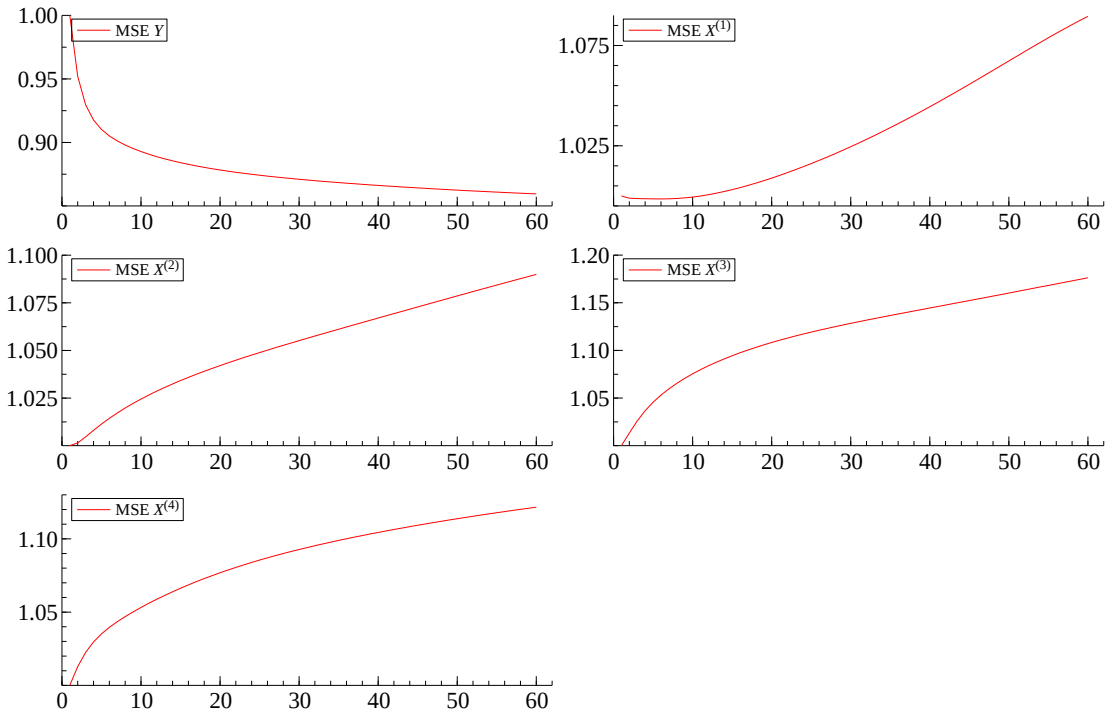
How much additive weight we give to y_t depends on the choice of W in (26). To investigate the effects of different values of W we study the in-sample fit. The MSE of the one-step ahead prediction errors of the Kalman filter is plotted as a function of W for y_t and for the four different series of x_t in Figure 2 for the MFS-M model. For each variable, all MSE values are divided by the MSE value at $W = 1$, in order to make the graphs comparable with each other.

By increasing W, the in-sample fit for y_t improves considerably, but the in-sample fits for all x_t become less accurate. Furthermore, from the first graph in Figure 2 it is clear that the marginal effect of increasing W becomes less for higher values of W. For example, increasing W from 1 to 6 leads to a 9.5% improvement in MSE for y_t while increasing W further to a value of 60 leads to a 14.0% improvement compared to $W = 1$. At $W = 6$ the largest increase in MSE is found for $x^{(3)}$ with 5.3%, while at $W = 60$ we find a 17.6% increase in MSE for this variable. Since the MSE of y_t is decreasing as a function of W, there is not one clear optimal value for W, but the optimum depends on how important the different variables are.

From the results presented in Figure 2 and from the results of our Monte Carlo study in Section 4.2, we may conclude that choosing the optimal value for W is a trade off between the gain in in-sample fit for y_t and its loss for x_t. However, when we focus on the effect of W on forecasting accuracy we obtain a different picture. In Table 6 we show the forecasting accuracy at different horizons for y_t for different values of W for the MFS-M model. It becomes apparent that increasing W also improves the accuracy of the forecasts for all horizons. Furthermore, we find that increasing the weight too much eventually leads to a decrease in the forecasting accuracy, because the fit of the x_t variables becomes worse with further increases in W. As a consequence, forecasted values for x_{t+h} become less informative for the forecasts of y_{t+h}. In our study we find that $W = 6$ is the optimal weight for $h = 0, 1, 2$ and $W = 5$ is optimal for $h = 3$. For $h = 6$ we find the lowest MSE with $W = 4$, so the optimal value for W appears to be lower for longer forecast horizons.

The Diebold and Mariano test indicates that the WML method with $W = 6$ leads to a significant improvement at a 5% level over the model with $W = 1$ (which provides the original estimates for the MFS-M model) for $h = 0, 1, 2$ and at a 10% level for $h = 3, 6$. The parameter estimates with $W = 5$ at $h = 3$ and $W = 4$ at $h = 6$ are also significant improvements over the models with $W = 1$ at these horizons.

Figure 2: In-sample accuracy using WML



This figure presents the in-sample MSE of the one-step ahead predictions for all variables in the Mixed frequency Stacking model with Monthly common factor (MFS-M) for different values of W in the weighted loglikelihood function (26). All models are estimated on the same dataset with sample period from January 1960 until December 2009 and using the data definitions as described in Table 3. The numbers on the horizontal axis represent the value of W and all numbers are relative to the values at $W = 1$.

In the next part of the study we take different integer values for W to determine the optimal value of W afterwards. Optimization techniques can be used to determine the optimal value of W for each forecasting horizon, as W is not limited to integer values only. However, our presented results provide a clear picture of the usefulness of the WML estimation method.

The weighted maximum likelihood method can be applied to all four dynamic factor models studied in this section. Based on our findings for the MFS-M model, we choose $W = 6$ as the optimal value and we estimate the MFI, MFS-Q and MFA models. The forecasting accuracy of the models with $W = 6$ is presented in Table 7. We find that for all models and for all forecasting horizons we obtain clear improvements when compared with the results for the models with $W = 1$ in Table 5. These improvements are all significant at 5% level. Furthermore, the MFS-M model has produced the most accurate forecasts for all horizons. Only in the case of $h = 6$ the MFI model performs best. For $h = 2$ and higher, all four models perform better than the benchmark BM and MIDAS models. For these benchmark models, weighted maximum likelihood is not possible, because these models are estimated using least squares methods.

Table 6: Forecast comparisons for US GDP growth with weighted likelihood function

W	$h=0$	$h=1$	$h=2$	$h=3$	$h=6$
1	0.1666	0.1730	0.2108	0.2935	0.3917
2	0.1600	0.1689	0.2049	0.2826	0.3708
3	0.1571	0.1674	0.2028	0.2783	0.3614
4	0.1556	0.1670	0.2013	0.2759	0.3534
5	0.1517	0.1703	0.2004	0.2745	0.3662
6	0.1513	0.1560	0.1914	0.2777	0.3733
7	0.1611	0.1668	0.2034	0.2773	0.3715
8	0.1608	0.1670	0.2033	0.2772	0.3699
9	0.1612	0.1682	0.2032	0.2775	0.3683
10	0.1614	0.1690	0.2033	0.2781	0.3662
11	0.1615	0.1698	0.2035	0.2786	0.3577
12	0.1617	0.1705	0.2037	0.2792	0.3572

This table presents the MSEs of the forecasts for the quarterly observed US GDP growth for the Mixed frequency Stacking model with Monthly common factor (MFS-M) at forecasting horizons $h = 0, 1, 2, 3, 6$ using different values for W in the weighted loglikelihood function (26). All models are estimated on the same dataset with sample period from January 1960 until December 2009 and using the data definitions as described in Table 3. For each forecasting horizon the MSE of the most accurate model is highlighted.

Table 7: Forecast comparisons for US GDP growth with WML and $W = 6$

	$h=0$	$h=1$	$h=2$	$h=3$	$h=6$
MFI	0.1787	0.1885	0.2078	0.2841	0.3629
MFS-M	0.1513	0.1560	0.1914	0.2777	0.3733
MFS-Q	0.1630	0.1676	0.2249	0.2849	0.3670
MFA	0.1576			0.2809	0.3677
BM	0.1833	0.2056	0.2455	0.3046	0.4197
MIDAS	0.1597	0.1658	0.2464	0.3635	0.4873

This table presents the MSEs of the forecasts for the quarterly observed US GDP growth for four different dynamic factor models and two benchmark models at forecasting horizons $h = 0, 1, 2, 3, 6$ using $W = 6$ in the weighted loglikelihood function (26) for the four dynamic factor models. The four dynamic factor models are Mixed frequency Interpolation (MFI), Mixed frequency Stacking with Monthly common factor (MFS-M), Mixed frequency Stacking with Quarterly common factor (MFS-Q) and Mixed frequency Aggregation (MFA). Benchmark models are Bridge models (BM) and MIDAS regressions. All models are estimated on the same dataset with sample period from January 1960 until December 2009 and using the data definitions as described in Table 3. For each forecasting horizon the MSEs of the two most accurate models are highlighted, where the darkest shade is used for the most accurate model.

However, the comparisons in Table 7 has given the MFS-M model some favor, because the value of $W = 6$ was chosen based on this model. Therefore, we next determine the optimal (integer) value of W for each model and for each forecasting horizon. These forecasting accuracy results are presented in Table 8. In the right panel of Table 8 we present the optimal values of W. The MFS-M model still performs best for all forecasting horizons, except for $h = 6$. Furthermore, it is very difficult to detect a clear pattern in the optimal values of W. Therefore, when adopting this method in a practical setting for

forecasting purposes, we would take a moderate fixed value for W, such as $W = 5$.

Table 8: Forecast comparisons for US GDP growth with WML and optimal value for W

	Mean Square Error				
	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 6$
MFI	0.1687	0.1765	0.1966	0.2835	0.3559
MFS-M	0.1513	0.1560	0.1914	0.2745	0.3593
MFS-Q	0.1629	0.1670	0.2215	0.2835	0.3621
MFA	0.1576			0.2769	0.3566
BM	0.1833	0.2056	0.2455	0.3046	0.4197
MIDAS	0.1597	0.1658	0.2464	0.3635	0.4873
	Optimal vale of W				
	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 6$
MFI	2	2	2	5	2
MFS-M	6	6	6	5	4
MFS-Q	7	8	3	3	2
MFA	6			2	8

The first panel of this table presents the MSEs of the forecasts for the quarterly observed US GDP growth for four different dynamic factor models and two benchmark models at forecasting horizons $h = 0, 1, 2, 3, 6$ using the optimal integer value for W in the weighted loglikelihood function (26) for the four dynamic factor models. The optimal values for W are presented in the second panel of the table. The four dynamic factor models are Mixed frequency Interpolation (MFI), Mixed frequency Stacking with Monthly common factor (MFS-M), Mixed frequency Stacking with Quarterly common factor (MFS-Q) and Mixed frequency Aggregation (MFA). Benchmark models are Bridge models (BM) and MIDAS regressions. All models are estimated on the same dataset with sample period from January 1960 until December 2009 and using the data definitions as described in Table 3. For each forecasting horizon the MSEs of the two most accurate models are highlighted, where the darkest shade is used for the most accurate model.

6 Conclusions

We have proposed solutions for two different problems that are often encountered in analyzing panels of economic and financial time series with different sample frequencies. The first problem is how to jointly model, say, monthly and quarterly time series. The default solution is to model both series as monthly variables and to allow for missing observations in the intermediate months for the quarterly variables. In our solution we opt for the modeling of the series in the lower quarterly sampling frequency and represent the monthly variable as a vector process to formulate the monthly dynamic process. There is no loss of high frequency information in our low frequency solution. We also show that in many situations, the low frequency solution leads to computational gains when compared to the high frequency solution. The second problem is how to focus on a set of key variables that we want to accurately forecast, in a simultaneous model. The typical setting is a dynamic factor analysis for a panel of time series with the purpose to forecast a small selection of variables in the panel. In our illustration it is a single series, the percentage growth in gross domestic product. We sacrifice some degree of fit for all other

variables with the aim to improve the fit of the variable of interest. Our solution consists of adding more weight to the likelihood contribution by the variable of interest in the construction of the likelihood function for the joint model. The resulting weighted likelihood function turns out to be rather effective in the accomplishment of targeting a specific variable of interest, whether it is for improving in-sample or out-of-sample performance of the model. We also show that the parameter estimates obtained from the maximization of the weighted likelihood are consistent and asymptotically normally distributed. Their small sample properties are investigated under different settings in a Monte Carlo study. The presented illustration shows that our solutions lead to empirically relevant improvements in nowcasting and forecasting. We expect that our proposed solutions also have consequences in other applications and in other modeling frameworks. Further research in these directions are planned in future.

A Derivations for Univariate Autoregressive Processes

A.1 AR(1) process with stacked observations

Consider the AR(1) process of the monthly (m) observed variable x_τ^m with monthly time index τ

$$x_\tau^m = \phi x_{\tau-1}^m + \varepsilon_\tau, \qquad \varepsilon_\tau \sim NID\left(0, \sigma_\varepsilon^2\right) \tag{30}$$

When the monthly observations of x_τ^m are stacked into the quarterly (q) 3×1 vectors x_t^q with quarterly time index t, then the equations of the AR(1) process for the stacked observations can be written as

$$x_{t,1}^q = \phi x_{t-1,3}^q + \varepsilon_{t,1}, \qquad x_{t,2}^q = \phi x_{t,1}^q + \varepsilon_{t,2}, \qquad x_{t,3}^q = \phi x_{t,2}^q + \varepsilon_{t,3} \tag{31}$$

To develop a low frequency recursion for x_t^q, we substitute the first equation for the value of $x_{t,1}^q$ and the second equation for $x_{t,2}^q$ gives the following set of equations

$$\begin{array}{rcl} x_{t,1}^q & = & \phi x_{t-1,3}^q + \varepsilon_{t,1} \\ x_{t,2}^q & = & \phi(\phi x_{t-1,3}^q + \varepsilon_{t,1}) + \varepsilon_{t,2} \\ & = & \phi^2 x_{t-1,3}^q + \phi\varepsilon_{t,1} + \varepsilon_{t,2} \\ x_{t,3}^q & = & \phi(\phi^2 x_{t-1,3}^q + \phi\varepsilon_{t,1} + \varepsilon_{t,2}) + \varepsilon_{t,3} \\ & = & \phi^3 x_{t-1,3}^q + \phi^2\varepsilon_{t,1} + \phi\varepsilon_{t,2} + \varepsilon_{t,3} \end{array}, \tag{32}$$

which can be written as the autoregressive process

$$x_t^q = T x_{t-1}^q + R\varepsilon_t \tag{33}$$

with matrices

$$T = \begin{pmatrix} 0 & 0 & \phi \\ 0 & 0 & \phi^2 \\ 0 & 0 & \phi^3 \end{pmatrix}, \qquad R = \begin{pmatrix} 1 & 0 & 0 \\ \phi & 1 & 0 \\ \phi^2 & \phi & 1 \end{pmatrix} \tag{34}$$

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

$$\begin{aligned} x_t &= Z\alpha_t + \varepsilon_t, & \varepsilon_t &\sim N(0, H) \\ \alpha_{t+1} &= T\alpha_t + R\varepsilon_t, & \varepsilon_t &\sim N(0, Q) \end{aligned} \tag{35}$$

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

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

$$\begin{aligned} \gamma_0 &= \tfrac{\sigma_\varepsilon^2}{(1-\phi^2)} \\ \gamma_1 &= \phi\gamma_0 \\ \gamma_2 &= \phi\gamma_1 \end{aligned} \tag{36}$$

A.2 AR(2) process with stacked observations

For an AR(2) process with monthly observed variable x_τ^m the transformations are similar. Consider the model

$$x_\tau^m = \phi_1 x_{\tau-1}^m + \phi_2 x_{\tau-2}^m + \varepsilon_\tau, \qquad \varepsilon_\tau \sim NID\left(0, \sigma_\varepsilon^2\right) \tag{37}$$

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

$$\begin{aligned} x_{t,1}^q &= \phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \varepsilon_{t,1} \\ x_{t,2}^q &= \phi_1 x_{t,1}^q + \phi_2 x_{t-1,3}^q + \varepsilon_{t,2} \\ x_{t,3}^q &= \phi_1 x_{t,2}^q + \phi_2 x_{t,1}^q + \varepsilon_{t,3} \end{aligned} \tag{38}$$

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

$$\begin{aligned} x_{t,1}^q &= \phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \varepsilon_{t,1} \\ x_{t,2}^q &= \phi_1(\phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,3}^q + \varepsilon_{t,1}) + \phi_2 x_{t-1,2}^q + \varepsilon_{t,2} \\ &= (\phi_1^2 + \phi_2) x_{t-1,3}^q + \phi_1\phi_2 x_{t-1,2}^q + \phi\varepsilon_{t,1} + \varepsilon_{t,2} \\ x_{t,3}^q &= \phi_1((\phi_1^2 + \phi_2) x_{t-1,3}^q + \phi_1\phi_2 x_{t-1,2}^q + \phi\varepsilon_{t,1} + \varepsilon_{t,2}) + \phi_2(\phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \varepsilon_{t,1}) + \varepsilon_{t,3} \\ &= (\phi_1^3 + 2\phi_1\phi_2) x_{t-1,3}^q + (\phi_1^2\phi_2 + \phi_2^2) x_{t-1,2}^q + (\phi_1^2 + \phi_2)\varepsilon_{t,1} + \phi_1\varepsilon_{t,2} + \varepsilon_{t,3} \end{aligned} \tag{39}$$

which can be written as the linear Gaussian state space model(35) with matrices

$$T = \begin{pmatrix} 0 & \phi_2 & \phi_1 \\ 0 & \phi_1\phi_2 & \phi_1^2+\phi_2 \\ 0 & \phi_1^2\phi_2+\phi_2^2 & \phi_1^3+2\phi_1\phi_2 \end{pmatrix}, \qquad R = \begin{pmatrix} 1 & 0 & 0 \\ \phi_1 & 1 & 0 \\ \phi_1^2+\phi_2 & \phi_1 & 1 \end{pmatrix} \tag{40}$$

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

$$\begin{aligned} \gamma_0 &= \tfrac{1-\phi_2}{1+\phi_2}\left(\tfrac{\sigma_\varepsilon^2}{(\phi_1+\phi_2-1)(\phi_2-\phi_1-1)}\right) \\ \gamma_1 &= \tfrac{\phi_1}{(1-\phi_2)}\gamma_0 \\ \gamma_2 &= \phi_1\gamma_1+\phi_2\gamma_0 \end{aligned} \tag{41}$$

A.3 AR(3) process with stacked observations

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

$$x_\tau^m = \phi_1 x_{\tau-1}^m + \phi_2 x_{\tau-2}^m + \phi_3 x_{\tau-3}^m + \varepsilon_\tau, \qquad \varepsilon_\tau \sim NID\left(0,\sigma_\varepsilon^2\right) \tag{42}$$

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

$$\begin{aligned} x_{t,1}^q &= \phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \phi_3 x_{t-1,1}^q + \varepsilon_{t,1} \\ x_{t,2}^q &= \phi_1 x_{t,1}^q + \phi_2 x_{t-1,3}^q + \phi_3 x_{t-1,2}^q + +\varepsilon_{t,2} \\ x_{t,3}^q &= \phi_1 x_{t,2}^q + \phi_2 x_{t,1}^q + \phi_3 x_{t-1,3}^q + +\varepsilon_{t,3} \end{aligned} \tag{43}$$

Substitutions similar to those described in the previous subsections can again be applied. We have

$$\begin{aligned} x_{t,1}^q &= \phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \phi_3 x_{t-1,1}^q + \varepsilon_{t,1} \\ x_{t,2}^q &= \phi_1 x_{t,1}^q + \phi_2 x_{t-1,3}^q + \phi_3 x_{t-1,2}^q + \varepsilon_{t,1} \\ &= \phi_1(\phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \phi_3 x_{t-1,1}^q + \varepsilon_{t,1}) + \phi_2 x_{t-1,3}^q + \phi_3 x_{t-1,2}^q + \varepsilon_{t,2} \\ &= (\phi_1^2+\phi_2) x_{t-1,3}^q + (\phi_1\phi_2+\phi_3) x_{t-1,2}^q + \phi_1\phi_3 x_{t-1,1}^q + \phi\varepsilon_{t,1} + \varepsilon_{t,2} \\ x_{t,3}^q &= \phi_1 x_{t,2}^q + \phi_2 x_{t,1}^q + \phi_3 x_{t-1,3}^q + \varepsilon_{t,1} \\ &= \phi_1(\phi_1 x_{t-1,3}^q + \phi_2 x_{t-1,2}^q + \phi_3 x_{t-1,1}^q + \varepsilon_{t,1}) \\ &+ \phi_2((\phi_1^2+\phi_2) x_{t-1,3}^q + (\phi_1\phi_2+\phi_3) x_{t-1,2}^q + \phi_1\phi_3 x_{t-1,1}^q + \phi\varepsilon_{t,1} + \varepsilon_{t,2}) \\ &+ \phi_3 x_{t-1,3}^q + \varepsilon_{t,2} \\ &= (\phi_1^3+2\phi_1\phi_2+\phi_3) x_{t-1,3}^q + (\phi_1^2\phi_2+\phi_1\phi_3+\phi_2^2) x_{t-1,2}^q \\ &+ (\phi_1^2\phi_3+\phi_2\phi_3) x_{t-1,1}^q + \phi\varepsilon_{t,1} + \varepsilon_{t,2} \end{aligned} \tag{44}$$

which can be written as the linear Gaussian state space model (35) for the quarterly observed x_t^q with

$$T = \begin{pmatrix} \phi_3 & \phi_2 & \phi_1 \\ \phi_1\phi_3 & \phi_1\phi_2 + \phi_3 & \phi_1^2 + \phi_2 \\ \phi_1^2\phi_3 + \phi_2\phi_3 & \phi_1^2\phi_2 + \phi_1\phi_3 + \phi_2^2 & \phi_1^3 + 2\phi_1\phi_2 + \phi_3 \end{pmatrix} \tag{45}$$

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

$$\begin{aligned} \gamma_0 &= \frac{\sigma_\varepsilon^2\left(1-\phi_2-\phi_1\phi_3-\phi_3^2\right)}{(1-\phi_2-\phi_3-\phi_1)\left(1+\phi_2+\phi_3\phi_1-\phi_3^2\right)(1+\phi_3+\phi_1-\phi_2)} \\ \gamma_1 &= \frac{(\phi_1+\phi_2\phi_3)\gamma_0}{\left(1-\phi_2-\phi_3\phi_1-\phi_3^2\right)} \\ \gamma_2 &= \phi_1\gamma_1 + \phi_2\gamma_0 + \phi_3\gamma_1 \end{aligned} \tag{46}$$

A.4 AR(p) process with stacked observations

For AR(p) processes of order $p > 3$ the state vector in (35) has to be extended with more lags of x_t^q. For example, for the AR(4) process we would have $\alpha_t = \left(x_{t-1,3}^q, x_{t,1}^q, x_{t,3}^q, x_{t,3}^q\right)'$ and the transition matrix becomes

$$T = \begin{pmatrix} 0 & 0 & 0 & 1 \\ \phi_4 & \phi_3 & \phi_2 & \phi_1 \\ \phi_1\phi_4 & \phi_1\phi_3 + \phi_4 & \phi_1\phi_2 + \phi_3 & \phi_1^2 + \phi_2 \\ \phi_1^2\phi_4 + \phi_2\phi_4 & \phi_1^2\phi_3 + \phi_2\phi_3 + \phi_1\phi_4 & \phi_1^2\phi_2 + \phi_1\phi_3 + \phi_2^2 + \phi_4 & \phi_1^3 + 2\phi_1\phi_2 + \phi_3 \end{pmatrix}, \tag{47}$$

where the variance matrix of the process is defined as $\sigma_\varepsilon^2 RR'$ with

$$R = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \phi_1 & 1 & 0 \\ 0 & \phi_1^2 + \phi_2 & \phi_1 & 1 \end{pmatrix}, \tag{48}$$

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

B Proofs of Theorems and Propositions

Proof of Proposition 1. Let $\{f_\tau(f_1^m)\}_{\tau\in\mathbb{N}}$ be generated according to (12) with initialization f_1^m and $\|T_f\| < 1$, $\|R_f\| < \infty$ and $\|\Sigma_\eta^2\| < \infty$. Then by Theorem 3.1 in Bougerol (1993), $\{f_\tau(f_1^m)\}_{t\in\mathbb{N}}$ converges to an SE sequence $\{f_\tau\}_{t\in\mathbb{Z}}$ satisfying $\mathbb{E}|f_\tau|^r < \infty \ \forall\ r$. Uniqueness of the limit SE sequence is obtained in Straumann and Mikosch (2006). Furthermore, since $\{f_t\}$ is a linear Gaussian process with $\mathbb{E}|x_\tau^m|^r < \infty \ \forall\ r > 0$. The bounds $|\beta_x| < \infty$ and $0 < \sigma_\epsilon^2 < \infty$, together with the iid Gaussian nature of the innovations $\{\epsilon_\tau\}^m$

ensure that $\{x_\tau^m\}$ is SE and Gaussian with bounded moments of any order. Similarly, the bounds $|\beta_y| < \infty$ and $0 < \sigma_\zeta^2 < \infty$ and the iid Gaussian nature of $\{\zeta_t\}$ ensure the SE linear Gaussian nature of $\{y_t\}$ with $\mathbb{E}|y_t|^r < \infty \; \forall \; r > 0$. □

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

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

$$\sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w, \tilde{f}_1^m) - \mathcal{L}_\infty(\psi, w)| \xrightarrow{a.s.} 0 \; \forall \; \tilde{f}_1^m \in \mathbb{R}_+ \;\; \text{as} \;\; T \to \infty, \tag{49}$$

and the identifiable uniqueness of the true parameter $\psi_0 \in \Psi$ defined e.g. in White (1994).

To establish the uniform convergence of $\mathcal{L}_T(\psi, w, \tilde{f}_1^m)$ we use the norm sub-additivity inequality

$$\sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w, \tilde{f}_1^m) - \mathcal{L}_\infty(\psi, w)| \;\leq\; \sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w, \tilde{f}_1^m) - \mathcal{L}_T(\psi, w)| + \sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w) - \mathcal{L}_\infty(\psi, w)|$$

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

$$\sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w, \tilde{f}_1^m) - \mathcal{L}_T(\psi, w)|$$

vanishes by the assumption that $\|\tilde{f}_t^m(\psi, w, \tilde{f}_1^m) - \tilde{f}_t^m(\psi)\| \xrightarrow{a.s.} 0$, the continuity of the likelihood function and the continuous mapping theorem.

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

$$\sup_{\psi \in \Psi} |\mathcal{L}_T(\psi, w) - \mathcal{L}_\infty(\psi, w)| \xrightarrow{a.s.} 0$$

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

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

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

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

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

is ensured by the uniqueness of ψ_0, the compactness of Ψ, and the continuity of $\mathbb{E}\ell_t(\psi,w)$ on Ψ, which is obtained through the continuity of $\mathcal{L}_T$ on Ψ for every $T\in\mathbb{N}$ and the uniform convergence of the likelihood; see e.g. White (1994). The uniqueness of ψ_0 as the maximizer $\mathcal{L}_\infty(\cdot,w)$ for any $w\in(0,1]$ is ensured by Theorem 3 which shows that the maximizer $\mathcal{L}_\infty(\psi_0,w)=0$ if and only if $\mathcal{L}_\infty(\psi_0,1)=0$. □

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

Let z_t denote a d_z-variate random vector with joint density $p(z_t)$. Furthermore, consider a family of parametric joint densities indexed by the parameter vector ψ, defined as $Q(\Psi):=\{q(z_t;\psi),\psi\in\Psi\}$. Note that it is possible but not necessary that $p(z_t)\in Q(\Psi)$. If $\psi_0^*(1)$ is the pseudo-true parameter that maximizes the limit log likelihood function

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

then it is well known that ψ_0^* also minimizes the the Kullback–Leibler divergence $\mathrm{KL}(p,q(\cdot;\psi))$ between $p(z_t)$ and $q(z_t;\psi)$ because

$$\arg\max_{\psi\in\Psi}\mathbb{E}_0\log q(z_t;\psi)=\arg\min_{\psi\in\Psi}\mathbb{E}_0\log p(z_t)-\mathbb{E}_0\log q(z_t;\psi)=\arg\min_{\psi\in\Psi}\mathrm{KL}\Big(p,q(\cdot;\psi)\Big).$$

Let now the joint density $q(x_t;\psi)$ be factorized into

$$q_1(z_{1,t}|z_{2:d_z,t};\psi_0)\times q_2(z_{2:d_z,t};\psi_0):=q_1(z_{1,t}|z_{2,t},...,z_{d_z,t};\psi_0)\times q_2(z_{2,t},z_{3,t},...,z_{d_z,t};\psi_0)$$

and define $\psi_0^*(w)$ as the pseudo-true parameter that maximizes the weighted limit log likelihood function

$$\psi_0^*(w):=\arg\max_{\psi\in\Psi}\mathbb{E}_0\log q_1(z_{1,t}|z_{2:d_z,t};\psi)+w\log\mathbb{E}_0 q_2(z_{2:d_z,t};\psi).$$

Then it follows naturally that $\psi_0^*(w)$ is the minimizer of the weighted average of KL divergences

$$\psi_0^*(w)=\arg\min_{\psi\in\Psi}\mathrm{KL}(q_1,p_1)+w\mathrm{KL}(q_2,p_2)$$

because

$$\begin{aligned}
&\arg\max_{\psi\in\Psi}\Big[\mathbb{E}_0\log q_1(z_{1,t}|z_{2:d_z,t};\psi)+w\log\mathbb{E}_0 q_2(z_{2:d_z,t};\psi)\Big]\\
&=\arg\max_{\psi\in\Psi}\Big[\mathbb{E}_0\log q_1(z_{1,t}|z_{2:d_z,t};\psi)-\mathbb{E}_0\log p_1(z_{1,t}|z_{2:d_z,t})\\
&\qquad\qquad+w\log\mathbb{E}_0 q_2(z_{2:d_z,t};\psi)-w\log\mathbb{E}_0 p_2(z_{2:d_z,t})\Big]\\
&=\arg\min_{\psi\in\Psi}\Bigg[\mathbb{E}_0\log p_1(z_{1,t}|z_{2:d_z,t})-\mathbb{E}_0\log q_1(z_{1,t}|z_{2:d_z,t};\psi)\\
&\qquad\qquad+w\Big(\log\mathbb{E}_0 p_2(z_{2:d_z,t})-\log\mathbb{E}_0 q_2(z_{2:d_z,t};\psi)\Big)\Bigg]\\
&=\arg\min_{\psi\in\Psi}\Big[\mathrm{KL}(q_1(\cdot;\psi),p_1)+w\mathrm{KL}(q_2(\cdot;\psi),p_2)\Big].
\end{aligned}$$

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

$$\psi_0^*(1)=\arg\min_{\psi\in\Psi}\mathrm{KL}(q_1(\cdot;\psi),p_1)+\mathrm{KL}(q_2(\cdot;\psi),p_2)=\arg\min_{\psi\in\Psi}\mathrm{KL}(q(\cdot;\psi),p).$$

For $w=0$ we obtain the ML estimator for the conditional model

$$\psi_0^*(0)=\arg\min_{\psi\in\Psi}\mathrm{KL}(q_1(\cdot;\psi),p_1)$$

Finally, it is also clear that the transformed KL divergence $\mathrm{TKL}(q,p):=\mathrm{KL}(q_1(\cdot;\psi),p_1)+w\mathrm{KL}(q_2(\cdot;\psi),p_2)$ satisfies

$$\mathrm{KL}(q_1(\cdot;\psi),p(\cdot;\psi))\leq\mathrm{TKL}(q(\cdot;\psi),p)\leq\mathrm{KL}(q(\cdot;\psi),p)$$

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

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

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

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

(i) $\hat{\psi}_T(\tilde{f}_1^m,w)\xrightarrow{a.s.}\psi_0\in\mathrm{int}(\Psi)$;

(ii) $\mathcal{L}_T(\cdot,w,\tilde{f}_1^m)\in\mathbb{C}^2(\Psi)$ a.s.;

(iii) $\sqrt{T}\mathcal{L}_T'(\psi_0,w,\tilde{f}_1^m,\tilde{df}_1^m)\xrightarrow{d}N(0,\mathcal{J}(\psi_0,w))$;

(iv) $\sup_{\psi\in\Psi}\left\|\mathcal{L}_T''(\psi,w,\tilde{f}_1^m,\tilde{df}_1^m,\tilde{dd}f_1^m)-\ell_\infty''(\psi,w)\right\|\xrightarrow{a.s.}0$;

(v) $\ell_\infty''(\psi,w)=\mathbb{E}\ell_t''(\psi,w)=\mathcal{I}(\psi,w)$ is non-singular

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

(iii) follows by the asymptotic SE nature of the score $\{\ell_t'(\psi, w, \tilde{f}_1^m, \tilde{df}_1^m)\}$ which is implied by the SE nature of $\{x_\tau^m\}$ and $\{y_t\}$, and the asymptotic SE nature of the filtered sequence $\{\tilde{f}_\tau(\psi, \tilde{f}_1^m)\}$ and its derivative $\{\tilde{df}_\tau(\psi, \tilde{df}_1^m)\}$. Since the limit score sequence $\{\ell_t'(\psi, w)\}$ is SE we can apply the CLT for SE martingales in Billingsley (1961) to obtain

$$\sqrt{T}\mathcal{L}_T'(\psi_0, w) \xrightarrow{d} N(0, \mathcal{J}(\psi_0, w)) \text{ as } T \to \infty, \tag{50}$$

where $\mathcal{J}(\psi_0, w) = \mathbb{E}(\ell_t'(\psi_0, w)\ell_t'(\psi_0, w)^\top) < \infty$. By Theorem 18.10[iv] in van der Vaart (2000) it thus follows that

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

as long as

$$\|\mathcal{L}_T'(\psi_0, w, \tilde{f}_1^m, \tilde{df}_1^m) - \mathcal{L}_T'(\psi_0, w)\| \xrightarrow{e.a.s.} 0 \text{ as } T \to \infty \tag{51}$$

since (51) ensures $\sqrt{T}\|\mathcal{L}_T'(\psi_0, w, \tilde{f}_1^m, \tilde{df}_1^m) - \mathcal{L}_T'(\psi_0, w)\| \xrightarrow{a.s.} 0$ as $T \to \infty$. The e.a.s. convergence in (51) follows from

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

and

$$\|\tilde{df}_\tau^m(\psi_0, \tilde{df}_1^m) - \tilde{df}_\tau^m(\psi_0)\| \xrightarrow{e.a.s.} 0.$$

Furthermore, since the score of the weighted likelihood is differentiable, we can use the mean-value theorem to obtain

$$\|\mathcal{L}_T'(\psi_0, w, \tilde{f}_1^m, \tilde{df}_1^m) - \mathcal{L}_T'(\psi_0, w)\| \leq \sum_j |d\mathcal{L}_T'||\tilde{df}_{j,\tau}^m(\psi_0, \tilde{df}_1^m) - \tilde{df}_{j,\tau}^m(\psi_0)|, \tag{52}$$

where $\tilde{df}_{j,\tau}^m(\psi_0, \tilde{df}_1^m)$ denotes the j-th element of $\tilde{df}_\tau^m(\psi_0, \tilde{df}_1^m)$, and $d\mathcal{L}_T'$ denotes the derivative $\partial\mathcal{L}_T'(\psi_0, w, \tilde{f}_1^m, \tilde{df}_1^m)/\partial\tilde{df}_j$ evaluated at some appropriate point between $\tilde{df}_{j,\tau}^m(\psi_0, \tilde{df}_1^m)$ and $\tilde{df}_{j,\tau}^m(\psi_0)$. The bounded moments of the weighted likelihood derivatives and the e.a.s. convergence of the filtered process and its derivatives yield

$$\|\mathcal{L}_T'(\psi_0, w, \tilde{f}_1^m, \tilde{df}_1^m) - \mathcal{L}_T'(\psi_0, w)\| = \sum_j O_p(1)o_{e.a.s}(1) = o_{e.a.s.}(1). \tag{53}$$

(iii) follows by noting that

$$\begin{aligned}\sup_{\psi\in\Psi} \|\mathcal{L}_T''(\psi, w, \tilde{f}_1^m, \tilde{df}_1^m, d\tilde{df}_1^m) - \mathcal{L}_\infty''(\psi, w) \leq \sup_{\psi\in\Psi} \|\mathcal{L}_T''(\psi, w, \tilde{f}_1^m, \tilde{df}_1^m, d\tilde{df}_1^m) - \mathcal{L}_T''(\psi, w)\| \\ + \sup_{\psi\in\Psi} \|\mathcal{L}_T''(\psi, w) - \mathcal{L}_\infty''(\psi, w)\|.\end{aligned} \tag{54}$$

Clearly, $\sup_{\psi\in\Psi}\|\mathcal{L}''_T(\psi,w,\tilde{f}_1^m,\tilde{df}_1^m,\tilde{ddf}_1^m)-\mathcal{L}''_T(\psi,w)\|0$ as $t\to\infty$ by the continuous mapping theorem and the e.a.s. convergence of the filtered process and its derivatives, and $\sup_{\psi\in\Psi}\|\mathcal{L}''_T(\psi,w)-\mathcal{L}''_\infty(\psi,w)\|$ vanishes by a ULLN under the uniform moment bound on the weighted likelihood $\mathbb{E}\sup_{\psi\in\Psi}\|\mathcal{L}''_t(\psi,w)\|<\infty$.

(v) the uniqueness of ψ_0 as a maximum of $\ell''_\infty(\psi,w)$ ensures the non-singularity of the limit weighted likelihood $\mathcal{L}''_\infty(\psi,w)=\mathbb{E}\mathcal{L}''_t(\psi,w)=\mathcal{I}(\psi,w)$. □

Proof of Theorem 5. Follows the same steps as the proof of Theorem 4 ith the exception that the required properties of $\{x_\tau^m\}$ and $\{y_t\}$ are directly assumed rather than derived through Proposition 1. □

C Technical Appendix: Recursive Algorithms

In general, observations from the monthly AR(p) processes of order p can be stacked into quarterly or yearly vectors. The system matrices T and R can both be obtained by recursive algorithms. We will first show how the recursions work for the AR(p) processes with $p=1,2,3$ that were described in the previous subsection. Then we will give the general steps of the recursion for AR(p) processes.

AR(1)

We will start with the example of the AR(1) process in (5) and (6). The transition matrix $T=[T_1',\ldots,T_s']'$, where T_i is the i-th row of matrix T for the low frequency autoregressive process in the state space representation, can be constructed as follows. From the equation $x_{t,1}^q=\phi x_{t-1,3}^q+\varepsilon_{t,1}$ we have $T_1=[0,0,\phi]$. For the subsequent rows we use the algorithm. Hence, we have $T_2=\phi T_1=\left[0,0,\phi^2\right]$ and $T_3=\phi T_2=\left[0,0,\phi^3\right]$. This may seem very trivial for the AR(1) process, but similar algorithms can be used for higher order AR(p) processes as well, as we will show in the next examples.

The matrix R in (6) can be constructed as follows. Start with ones on the diagonal and zeros for all upper-diagonal elements. We refer to the diagonal elements as R_D. The elements that are one position under the diagonal are denoted as R_{D+1} and are equal to $\phi R_D=\phi$. For the elements (in this case only one) two positions under the diagonal we have $R_{D+2}=\phi R_{D+1}=\phi^2$.

AR(2)

For the AR(2) process the matrix T in (7), with $\phi_3=0$, can also be constructed with the recursive algorithm. From the equation $x_{t,1}^q=\phi_1 x_{t-1,3}^q+\phi_2 x_{t-1,2}^q+\varepsilon_{t,1}$ we have $T_1=[0,\phi_2,\phi_1]$. To construct the second row T_2 we start with the equation $x_{t,2}^q=\phi_1 x_{t,1}^q+\phi_2 x_{t-1,3}^q+\varepsilon_{t,2}$ and substitute the first equation for the value of $x_{t,1}^q$. This gives us

$$\begin{aligned} x^q_{t,2} &= \phi_1(\phi_1 x^q_{t-1,3} + \phi_2 x^q_{t-1,2} + \varepsilon_{t,1}) + \phi_2 x^q_{t-1,2} + \varepsilon_{t,2} \\ &= (\phi_1^2 + \phi_2) x^q_{t-1,3} + \phi_1\phi_2 x^q_{t-1,2} + \phi\varepsilon_{t,1} + \varepsilon_{t,2} \end{aligned} , \tag{55}$$

which gives us $T_2 = [0, (\phi_1\phi_2), (\phi_1^2 + \phi_2)]$. The third row T_3 can now be constructed using the algorithm $T_j = \phi_1 T_{j-1} + \phi_2 T_{j-2}$. In this example, the recursive formula only has to be used for $j = 3$.

The matrix R can be constructed as follows. Start with ones on the diagonal and zeros for all upper-diagonal elements. We again refer to the diagonal elements as R_D. For the elements under the diagonal we have $R_{D+j} = \phi_1 R_{D+j-1} + \phi_2 R_{D+j-2}$.

AR(3)

For the AR(3) process, matrix R is the same matrix as for the AR(2) process, which can be explained as follows. The elements under the unit diagonal of matrix R are given by the equation $R_{D+j} = \phi_1 R_{D+j-1} + \phi_2 R_{D+j-2} + \phi_3 R_{D+j-3}$. However, because in this example R is only a 3×3 matrix, the term $\phi_3 R_{D+j-3}$ is never effectively used.

In this example the elements of all the three rows of T in (7) have to be derived manually. From the equation $x^q_{t,1} = \phi_1 x^q_{t-1,3} + \phi_2 x^q_{t-1,2} + \phi_3 x^q_{t-1,1} + \varepsilon_{t,1}$ we have $T_1 = [\phi_3, \phi_2, \phi_1]$. Substitutions as in (55) are needed to construct T_2 and T_3. We have

$$\begin{aligned} x^q_{t,2} &= \phi_1 x^q_{t,1} + \phi_2 x^q_{t-1,3} + \phi_3 x^q_{t-1,2} + \varepsilon_{t,1} \\ &= \phi_1(\phi_1 x^q_{t-1,3} + \phi_2 x^q_{t-1,2} + \phi_3 x^q_{t-1,1} + \varepsilon_{t,1}) + \phi_2 x^q_{t-1,3} + \phi_3 x^q_{t-1,2} + \varepsilon_{t,2} \\ &= (\phi_1^2 + \phi_2) x^q_{t-1,3} + (\phi_1\phi_2 + \phi_3) x^q_{t-1,2} + \phi_1\phi_3 x^q_{t-1,1} + \phi\varepsilon_{t,1} + \varepsilon_{t,2} \end{aligned} , \tag{56}$$

which gives us $T_2 = [\phi_1\phi_3, (\phi_1\phi_2 + \phi_3), (\phi_1^2 + \phi_2)]$. The third row T_3 is constructed as follows

$$\begin{aligned} x^q_{t,3} &= \phi_1 x^q_{t,2} + \phi_2 x^q_{t,1} + \phi_3 x^q_{t-1,3} + \varepsilon_{t,1} \\ &= \phi_1(\phi_1 x^q_{t-1,3} + \phi_2 x^q_{t-1,2} + \phi_3 x^q_{t-1,1} + \varepsilon_{t,1}) \\ &+ \phi_2((\phi_1^2 + \phi_2) x^q_{t-1,3} + (\phi_1\phi_2 + \phi_3) x^q_{t-1,2} + \phi_1\phi_3 x^q_{t-1,1} + \phi\varepsilon_{t,1} + \varepsilon_{t,2}) \\ &+ \phi_3 x^q_{t-1,3} + \varepsilon_{t,2} \\ &= (\phi_1^3 + 2\phi_1\phi_2 + \phi_3) x^q_{t-1,3} + (\phi_1^2\phi_2 + \phi_1\phi_3 + \phi_2^2) x^q_{t-1,2} \\ &+ (\phi_1^2\phi_3 + \phi_2\phi_3) x^q_{t-1,1} + \phi\varepsilon_{t,1} + \varepsilon_{t,2} \end{aligned} , \tag{57}$$

which gives us $T_3 = [(\phi_1^2\phi_3 + \phi_2\phi_3), (\phi_1^2\phi_2 + \phi_1\phi_3 + \phi_2^2), (\phi_1^3 + 2\phi_1\phi_2 + \phi_3)]$.

The elements of subsequent rows would be given by the equation $T_j = \phi_1 T_{j-1} + \phi_2 T_{j-2} + \phi_3 T_{j-3}$. However, we do not need the recursion in this example, since the matrix T only has 3 rows.

AR(3) WITH YEARLY FREQUENCY

This becomes different when the monthly observations of x^m_τ are stacked into a 12×1 vector with yearly observations x^y_t. The equations for the AR(3) process then become

$$\begin{aligned} x^y_{t,1} &= \phi_1 x^y_{t-1,12} + \phi_2 x^y_{t-1,11} + \phi_3 x^y_{t-1,10} + \varepsilon^y_{t,1} \\ x^y_{t,2} &= \phi_1 x^y_{t,1} + \phi_2 x^y_{t-1,12} + \phi_3 x^y_{t-1,11} + \varepsilon^y_{t,2} \\ &\vdots \\ x^y_{t,12} &= \phi_1 x^y_{t,11} + \phi_2 x^y_{t,10} + \phi_3 x^y_{t,9} + \varepsilon^y_{t,12} \end{aligned} \tag{58}$$

The 12×12 matrix T can again be constructed using the equation $T_j = \phi_1 T_{j-1} + \phi_2 T_{j-2} + \phi_3 T_{j-3}$, for $j = 4, \ldots, 12$ and with

$$\begin{pmatrix} T_1 \\ T_2 \\ T_3 \end{pmatrix} = \begin{pmatrix} 0 & \ldots & 0 & \phi_3 & \phi_2 & \phi_1 \\ 0 & \ldots & 0 & \phi_1\phi_3 & \phi_1\phi_2 + \phi_3 & \phi_1^2 + \phi_2 \\ 0 & \ldots & 0 & \phi_1^2\phi_3 + \phi_2\phi_3 & \phi_1^2\phi_2 + \phi_1\phi_3 + \phi_2^2 & \phi_1^3 + 2\phi_1\phi_2 + \phi_3 \end{pmatrix} \tag{59}$$

where the last three columns are equal to the matrix T of the quarterly AR(3) process in (7).

The matrix R is again constructed using $R_{D+j} = \phi_1 R_{D+j-1} + \phi_2 R_{D+j-2} + \phi_3 R_{D+j-3}$ for $j = 1, \ldots, 11$ and starting with the 12×12 identity matrix I_{12}.

AR(p)

To summarize, for monthly AR(p) processes of order p that are stacked into quarterly or yearly vectors the transition matrix T of the autoregressive process and state space representation is obtained by the following recursive algorithm:

- Construct the first p rows of T by rewriting the equations for the first p elements of x^q_t in terms of elements of x^q_{t-1}.
- When the vector length s is larger than p, calculate subsequent rows using $T_j = \phi_1 T_{j-1} + \ldots + \phi_p T_{j-p}$ for $j = p+1, \ldots, s$ (with $s = 3$ for quarterly and $s = 12$ for yearly vectors).

The matrix R is obtained by the following recursive algorithm:

- Start with the identity matrix I_3 (quarterly) or I_{12} (yearly) and refer to the diagonal elements as R_D.
- Elements j positions under the diagonal are referred to as R_{D+j}. Calculate these elements as $R_{D+j} = \phi_1 R_{D+j-1} + \ldots + \phi_p R_{D+j-p}$ for $j = 1, 2$ (quarterly) or $j = 1, \ldots, 11$ (yearly).

Note that when the order p of the AR(p) process exceeds the vector length s, earlier lags have to be added to the state vector. See Section 2.3 for an example with an AR(4) process and 3×1 vectors of observations.

D Technical Appendix: Benchmark Models

D.1 Bridge Models

Bridge Models are linear dynamic equations where the aggregate GDP is explained by suitable short-term indicators. For this comparison we use the same four monthly indicators (log-transformed and demeaned) as in the dynamic factor models described in the previous subsection. The monthly indicators are averaged in quarterly series, such that they have the same frequency as GDP and a linear regression model with GDP as the dependent variable is estimated. For this comparison, we also include two lags of GDP in the equation

$$y_t = \beta_{y,1} y_{t-1} + \beta_{y,2} y_{t-2} + \beta_1 x_t^{(1)} + \beta_2 x_t^{(2)} + \beta_3 x_t^{(3)} + \beta_4 x_t^{(4)} + \varepsilon_t \tag{60}$$

where all variables are log-transformed and demeaned, so no constant is needed in the equation.

Forecasting with BM is done in two steps: first the monthly indicators are forecasted up to the desired forecasting horizon, using separate AR(2) models for the four different indicators. For example, when forecasting the next quarter GDP value $y_{(t+1)}$ and the conditioning indicators are already known for the first month of the quarter $(t+1)$, the indicators have to be forecast two months ahead by the AR(2) models. In the second step, the forecasts and known values are aggregated to quarterly averages again and plugged into the estimated regression equation, together with the lagged values of GDP. All parameters of the model can be estimated with Least Squares.

D.2 MIDAS Regressions

The MIDAS approach was proposed by Ghysels, Santa-Clara, and Valkanov (2006). Similar to in BM, the series with the higher frequency are regressed on the series with the lower frequency. However, when forecasting with MIDAS regression only one step is required. To obtain forecasts at horizon h, the values of y_t are simply regressed on the values of the indicators up to period $t-h$ and the dynamics of the regressors are not specified by the model.

In MIDAS regressions, the higher frequency series are not aggregated, but each lag has its own regression coefficient. This means that in our example, three regression coefficients have to be estimated per indicator. To avoid parameter proliferation, the coefficients of the different lags are described by a weighting function w_j, such that the regression equation for this example becomes

$$y_{t+h} = \beta_y \sum_{j=1}^{2} w_j(\theta_y) y_{t-j} + \sum_{i=1}^{4} \beta_{x^{(i)}} \sum_{j=1}^{6} w_j(\theta_{x^{(i)}}) x_{t-j/m}^{(i)} + \varepsilon_{t+h} \tag{61}$$

where $m = 3$, because each quarter consists of three months. Again all variables are log-transformed and demeaned, so no constant is needed in the equation. Two lags of y are

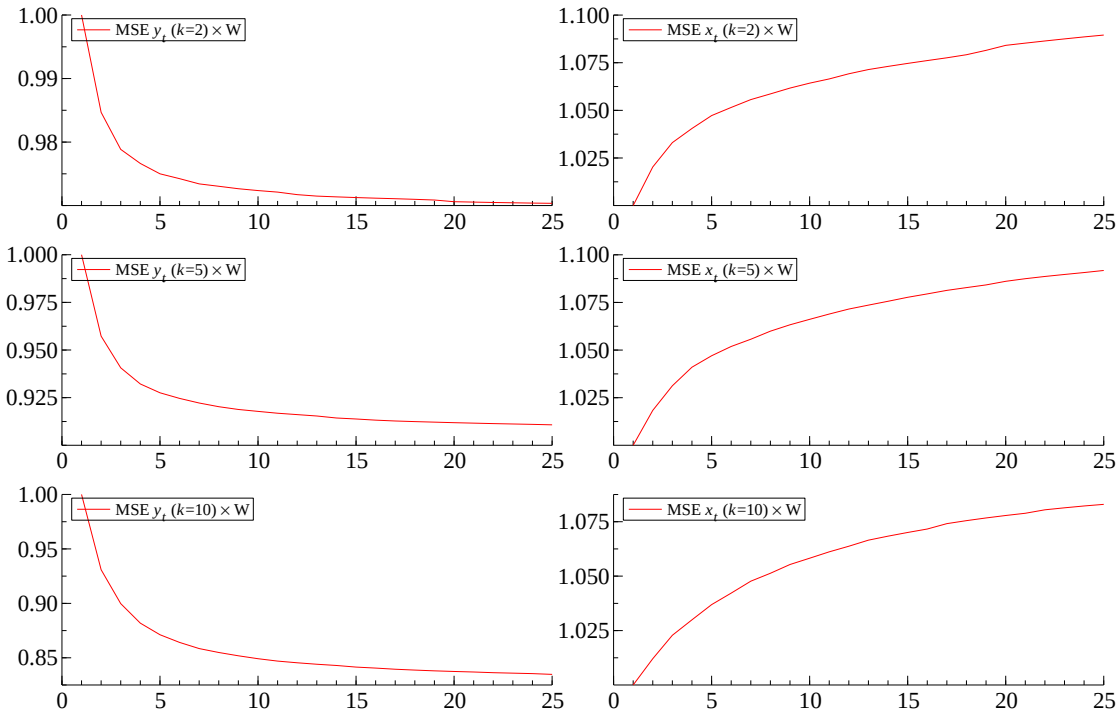
included in the equation, and six lags (e.g. two quarters) for each x_i. These numbers were chosen in order to make all models under investigation in this comparison use roughly the same amount of information in forecasting. The weighting schemes $w_j(\theta_y)$ and $w_j(\theta_{x_i})$ are two-parameter exponential Almon lag polynomials:

$$w_j(\theta_1, \theta_2) = \frac{\exp(\theta_1 j + \theta_2 j^2)}{\sum_{j=1}^{6} \exp(\theta_1 j + \theta_2 j^2)} \tag{62}$$

Note that the weights are governed by two parameters and scaled such that they add up to one. The parameters can be estimated with Least Squares. When the model is correctly specified and the parameters are known, the Kalman filter is superior to MIDAS by construction. Otherwise, it is under investigation whether MIDAS or the state space method is superior; see the study of Bai, Ghysels, and Wright (2011) where both MIDAS and state space methods are considered. They show under which conditions the methods are identical and provide evidence that the Kalman filter is slightly more accurate.

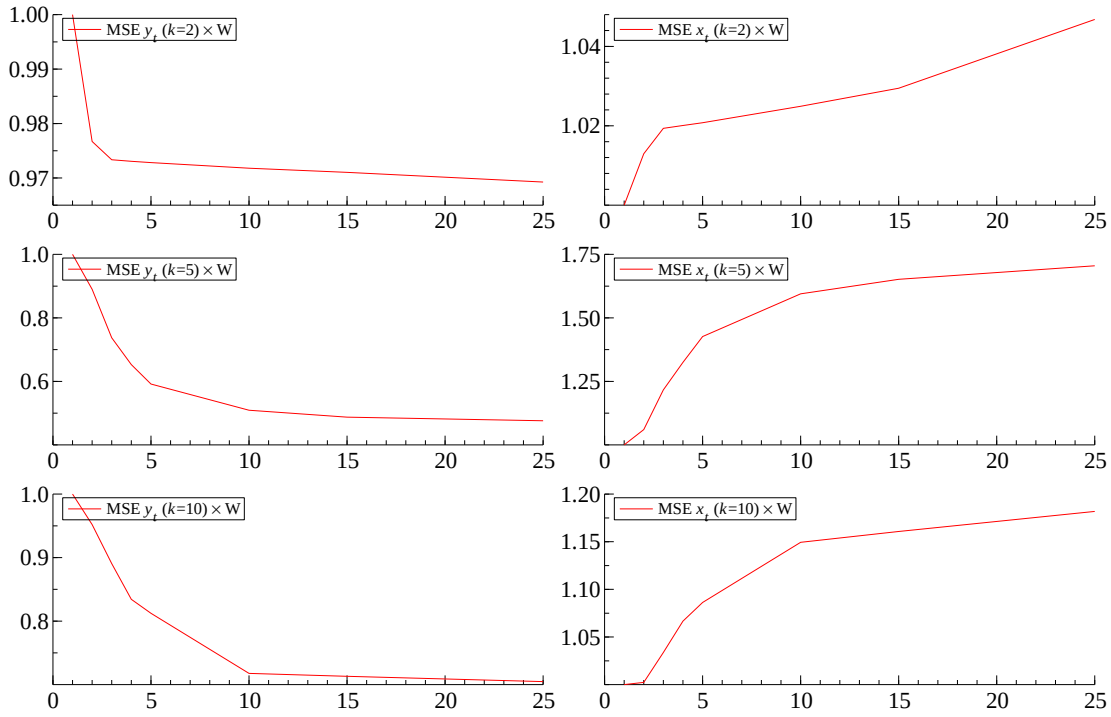
E Technical Appendix: Weighted Maximum Likelihood: MSE Figures

Figure 3: Scenario 1: Underspecification



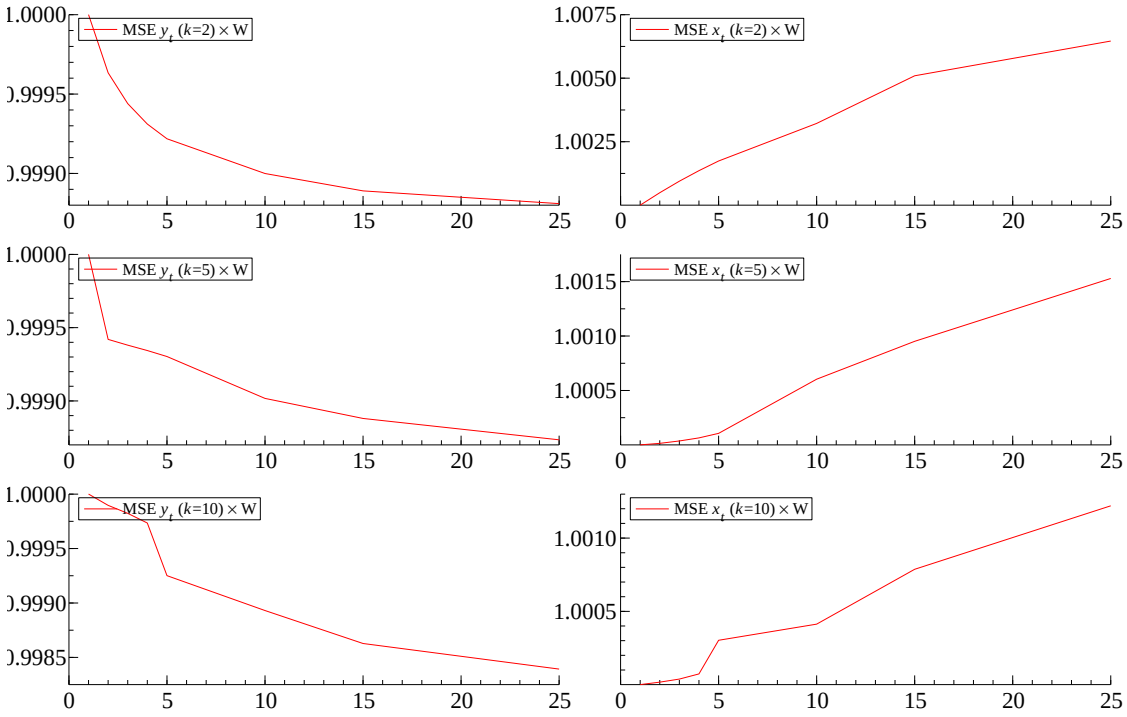
This figure presents the in-sample MSE of the one-step ahead predictions using an underspecified model for different values of W in the weighted loglikelihood function (26) for $k = 2$ (upper panels), $k = 5$ (middle panels) and $k = 10$ (lower panels) variables. The DGP is a common factor model with idiosyncratic factors and the estimated model is a model with only a common factor. The numbers on the horizontal axis represent the value of W and all numbers are relative to the values at $W = 1$. The MSEs of the target variable y_t are presented in the figures on the right. The panels on the left present the average MSEs of the other variables x_t.

Figure 4: Scenario 2: Misspecification



This figure presents the in-sample MSE of the one-step ahead predictions using a misspecified model for different values of W in the weighted loglikelihood function (26) for $k = 2$ (upper panels), $k = 5$ (middle panels) and $k = 10$ (lower panels) variables. The DGP is a VAR(1) model and the estimated model is a dynamic factor model with one common factor. The numbers on the horizontal axis represent the value of W and all numbers are relative to the values at $W = 1$. The MSEs of the target variable y_t are presented in the figures on the right. The panels on the left present the average MSEs of the other variables x_t.

Figure 5: Scenario 3: Correct specification



This figure presents the in-sample MSE of the one-step ahead predictions using a correctly specified model for different values of W in the weighted loglikelihood function (26) for $k = 2$ (upper panels), $k = 5$ (middle panels) and $k = 10$ (lower panels) variables. The DGP and the estimated model are both a common factor model with idiosyncratic factors. The numbers on the horizontal axis represent the value of W and all numbers are relative to the values at $W = 1$. The MSEs of the target variable y_t are presented in the figures on the right. The panels on the left present the average MSEs of the other variables x_t.

References

Aruoba, S., F. X. Diebold, and C. Scotti (2008). Real-time measurement of business conditions, second version. *PIER Working Paper Archive No. 11, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania*. Working paper.

Baffigi, A., R. Golinelli, and G. Parigi (2004). Bridge models to forecast the euro area gdp. *International Journal of Forecasting 20*, 447–460.

Bai, J., E. Ghysels, and J. H. Wright (2011). State space models and midas regressions. Working paper.

Banbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Nowcasting and the real time data flow. *Handbook of Economic Forecasting 2A*.

Billingsley, P. (1961). The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society 12*(5), 788–792.

Blasques, F., S. J. Koopman, and A. Lucas (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 Journal on Control and Optimization 31*(4), 942–959.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *JBES 13*(3), 253–63.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Foroni, C., M. Marcellino, and C. Schumacher (2012). U-midas: Midas regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society*. Forthcoming.

Gallant, R. and H. White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Cambridge University Press.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics 131*, 59–95.

Golinelli, R. G. and Parigi (2007). The use of monthly indicators to forecast quarterly gdp in the short-run: An application to the g7 countries. *Journal of Forecasting 26*, 77–94.

Koopman, S. J. and J. Durbin (2000). Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis 21*, 281–96.

Krengel, U. (1985). *Ergodic theorems*. Berlin: De Gruyter studies in Mathematics.

Marcellino, M., A. Carriero, and T. E. Clark (2014). Real-time nowcasting with a bayesian mixed frequency model with stochastic volatility. Working paper.

Mariano, R. and Y. Murasawa (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics 18*, 427–443.

Mehra, R. (1970). On the identification of variances and adaptive kalman filtering. *IEEE Transactions on Automatic Control 15*(2).

Rao, R. R. (1962). Relations between Weak and Uniform Convergence of Measures with Applications. *The Annals of Mathematical Statistics 33*(2), 659–680.

Stock, J. H. and M. Watson (1990). A probability model of the coincident economic indicators. In K. Lahiri and G. H. Moore (Eds.), *Leading Economic Indicators*, pp. 63–89. Cambridge: Cambridge University Press.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroeskedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics 34*(5), 2449–2495.

Trehan, B. (1989). "forecasting growth in current quarter real gnp. *Federal Reserve Bank of S. Francisco Economic Review Winter*, 39–52.

van der Vaart, A. W. (2000, June). *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge Books. Cambridge University Press.

Wohlrabe, K. (2009). Forecasting with mixed-frequency time series models. *Munich Dissertations in Economics 9681*.