Journal of Forecasting

WILEY

RESEARCH ARTICLE OPEN ACCESS

Forecasting With Dynamic Factor Models Estimated by Partial Least Squares

Samuel Rauhala

Department of Mathematics and Statistics, University of Turku, Turku, Finland

Correspondence: Samuel Rauhala (samuel.j.rauhala@utu.fi)

Received: 17 September 2025 | **Revised:** 19 February 2026 | **Accepted:** 2 April 2026

Keywords: dynamic factor model | forecasting | GDP growth | high-dimensional time series | partial least squares

ABSTRACT

Dynamic factor models (DFMs) have found great success in nowcasting and short-term macroeconomic forecasting when incorporating large sets of predictive information. The factor loadings are typically estimated cross-sectionally with principal component analysis (PCA) or maximum likelihood (ML), which ignore whether the factors have predictive power. We suggest two novel alternative approaches using partial least squares to estimate large vector autoregressions (VARs) and DFMs, which take the dynamic dependencies better into account. Our Monte Carlo simulations and forecasting results for the Finnish GDP growth show that these methods generally perform on par with and under certain conditions better than the existing approaches.

1 | Introduction

A common approach to accommodate big data in macroeconometrics is to use dynamic factor models (DFMs). These models rely on the fact that macroeconomic data have strong comovements that can often be reduced to variations in just a few underlying factors. This dimension reduction step makes it possible to model huge panels of data without having to estimate too many parameters.

Over time, multiple ways to estimate DFMs have been introduced: First studies employed complex frequency-domain methods (Geweke 1977; Sargent and Sims 1977). It was soon found easier to express DFMs in a state-space form and to estimate the model using the method of maximum likelihood (ML) (Engle and Watson 1981; 1983; Stock and Watson 1989). However, ML estimation struggled with high-dimensional data and principal component analysis (PCA) has often been used instead to estimate the models. While PCA is consistent and computationally lighter, it is less efficient than ML and struggles with parameter restrictions which are used with mixed frequency data (Chamberlain and Rothschild 1983; Connor and Korajczyk 1986; Forni and Reichlin 1996; 1998; Doz et al. 2011). More recently, efficient and easily applicable approaches have been developed using maximum likelihood (ML) or quasi-ML (QML) based estimation (see Doz et al. 2012; Bańbura and Modugno 2014; Jungbacker and Koopman 2015).

In this study, we consider an approach that has received far less attention: partial least squares (PLS). Whereas conventional methods, like PCA and ML estimation, find the rotations of the predictor variable matrix X that best explain the variation in it, PLS finds rotations that also explain variations in the response (matrix) Y. In the context of the vector autoregression, X contains the lags of Y. Because the rotations are done at the same time for both the response and its lags, these rotations can be used to either construct a vector autoregression (VAR) or a DFM without a need to separately estimate the transition matrix or a similar auxiliary quantity.

When estimating a DFM, PLS has two main advantages over PCA or ML. First, by the construction of factors, it also uses information in the lags of the endogenous variable. This increases the width of the data used, which is useful since factor methods generally benefit from increasing both the size of the cross-section and the number of observations. Second, if it is

infeasible to estimate all the factors, PLS can be used to estimate only those factors that have predictive power. This allows for more parsimonious models. Moreover, and importantly from a practical perspective, PLS is also computationally lighter than the burdensome EM-algorithm used in ML estimation.

The main contribution of this work is to develop PLS to estimate a DFM. This is achieved by a variation of the standard PLS algorithm of Höskuldsson (1988). Our approach is well suited for high-dimensional mixed-frequency time series data with missing observations, which is generally a highly relevant setup for various macroeconomic applications related to nowcasting and forecasting. The same algorithm is also applicable for large vector autoregressions (VARs). Our Monte Carlo simulation results suggest that the PLS based approach can perform equally or better than the conventional ML estimation, when the data is generated by a DFM. Furthermore, empirical nowcasting and forecasting results, obtained with a large panel of Finnish macroeconomic data, also suggest good performance by the PLS-based methods in nowcasting and forecasting quarterly Finnish GDP growth.

The rest of the paper is organized as follows. Primers of both DFMs and PLS are presented in Section 2. The formulation and estimation of the PLS based VARs (VAR-PLSR) and DFMs (PLS-DFM) is given in Section 3. Simulations and empirical results are presented in Sections 4 and 5, respectively. Finally, Section 6 concludes.

2 | Modeling High-Dimensional Data

2.1 | Dynamic Factor Models (DFM)

Dynamic Factor Models (DFMs) have been found to perform well in nowcasting and short-term forecasting of macroeconomic time series, such as and especially Gross Domestic Product (GDP) (see, for example, Hindrayanto et al. 2016 and Stundziene et al. 2024 for a literature review). A simple DFM is given by

$$y_t = \Lambda f_t + \xi_t, \tag{1a}$$

$$f_t = \beta f_{t-1} + w_t,\ t = 1,\ \dots, T, \tag{1b}$$

where y_t is a n-dimensional vector of demeaned observed variables, f_t is a q-dimensional vector of latent factors with $q \ll n$, ξ_t is n-dimensional idiosyncratic error term, and w_t is a q-dimensional vector of innovations. In the exact DFM, the covariance matrix of ξ_t is diagonal, and ξ_t and y_t are uncorrelated with ξ_{t+h} at every $h \neq 0$, while in the approximate model, non-trivial correlation structures and especially autocorrelation in ξ_t, are allowed. The parameters in (1a–1b) are contained in an $n \times q$ matrix Λ and a $q \times q$ matrix β. It is straightforward to extend model (1a–1b) by increasing the number of lags in either equation: y_t may contain the lagged values of f_t, and f_t can also depend on additional lags of itself. For further discussion, see Stock and Watson (2016).

From (1a), it can be seen that $\mathrm{Cov}(y_t) = \Lambda \mathrm{Cov}(f_t)\Lambda' + \mathrm{Cov}(\xi_t)$. Factor models often make additional assumptions about these covariance matrices. The key assumption is that the first term, $\Lambda \mathrm{Cov}(f_t)\Lambda'$, captures most of the co-movements so that $\mathrm{Cov}(\xi_t)$ is at least nearly diagonal. In other words, the idiosyncratic component ξ_t cannot have a factor structure of its own, which in turn sets assumptions about the maximum eigenvalues of $\mathrm{Cov}(\xi_t)$ as $n \to \infty$. For additional discussion of these assumptions, see Doz et al. (2011, 2012).

The earliest papers by Geweke (1977) and Sargent and Sims (1977) estimate the DFM using frequency-domain methods. However, as this proved complex and laborious, the method of maximum likelihood has become more commonplace. Because the factors f_t are not observed, they must be estimated as part of the estimation process, which calls for the Expectation Maximization (EM) algorithm. Typically ξ_t and w_t are assumed to follow a multivariate normal distribution and if no restrictions are imposed on their covariances, the model becomes heavily parameterized, which impedes ML estimation. As a consequence, early maximum likelihood-based literature either made strong assumptions about these distributions or only dealt with a few time series (see, for example, Engle and Watson 1981, 1983; Stock and Watson 1989). Additionally, the early literature struggled with computational limitations, as the EM-algorithm requires many iterations and the estimation of factors over the whole sample.

From an alternative perspective, the DFM can be seen as analogous to a specific principal component regression (PCR). In a simple PCR framework, the principal components of a set of predictors, in our case lagged values of y_t, are used to predict a single variable, y_{it}. This would correspond to a model

$$y_{it} = \beta_i^{\mathrm{pcr}} f_t^{\mathrm{pc}} + \xi_{it}^{\mathrm{pcr}}, \tag{2}$$

where f_t^{pc} are principal components (q-dimensional) of lagged values of y_t and β_i^{pcr} are estimated by ordinary least squares (OLS) on f_t^{pc}. These principal components, f_t^{pc}, can be seen as estimates of factors in a very simple factor model. Equation (2), when written for all values of $i = 1,\ \dots, n$ makes up (1a).

To make the model dynamic, the principal components can be modeled with a VAR like in (1b). PCA was first used in the DFM context by Chamberlain and Rothschild (1983), and later it was found that PCA is a consistent way to estimate the factors (see, for example, Connor and Korajczyk 1986; Forni and Reichlin 1996, 1998). PCA factor estimates can be further improved with the Kalman filter, and to that end, Doz et al. (2011) recommend a two-step approach (2s-DFM). In the first step, Λ is estimated by PCA on y_t, and β is then estimated by OLS on the principal components of that PCA. The covariances of ξ_t and w_t are estimated on the residuals of y_t and the principal components, respectively. Then in the second stage, f_t is estimated by a Kalman filter using the estimates for various parameters from the first step.[1]

In the more recent literature, the maximum likelihood approach has become the predominant approach largely due to advances in computing. Doz et al. (2012) use a (QML) estimator to estimate an exact DFM (non-autocorrelated ξ_t) and show that it is, in fact, consistent and more efficient than the two-step estimator even when the true model is approximate (autocorrelated ξ_t). Bańbura and Modugno (2014) provide an algorithm to estimate the model with autocorrelated idiosyncratic components. Once cross-correlation between the components of ξ_t is ignored, the

Kalman filter makes this easy when we include the idiosyncratic components in the state variable with the factors. Others, like Jungbacker and Koopman (2015), have developed computationally efficient ML estimation procedures.

2.2 | Partial Least Squares Regression (PLSR)

PLS estimation has been extensively studied from various perspectives, but this paper is only concerned with PLS regression (hereafter PLSR). Furthermore, this paper uses language typical of the econometrics literature and thus the terminology may slightly differ from other papers on PLS.

A common criticism of the PCR of Equation (2) is that as the factors are estimated to explain the cross-sectional variation, the number of factors can be inflated by factors that do not help in forecasting even if PCA is able to find all relevant factors. This leads to needlessly complex models, which increases estimation variance. Helland (1990) gives a more formal motivation for the PCR and the PLSR: The idea behind using principal components as predictors is that we can often assume that some eigenvalues of the predictors' covariance matrix are zero on the population level. This means that the predictor space is spanned by a smaller number of vectors than the dimension of the predictors. Therefore, the same predictive power can be achieved with fewer predictors when principal components are used instead. Meanwhile, the idea behind PLS regression is that in addition to ignoring the eigenvectors that correspond to eigenvalues that are zero, also the eigenvectors that have zero covariance with the response are ignored, which typically leads to significantly more parsimonious models. For our purposes, it is important to note that DFMs estimated with ML share this same weakness with PCR.

The general structure of a PLSR used in this paper is expressed as

$$y_t = \Lambda_y D \Lambda_x' x_t + \xi_t, \ t = 1, \ \dots, T, \tag{3}$$

where y_t remains the n-dimensional response vector, x_t is a k-dimensional vector of predictors (later in Section 3 lags of y_t), Λ_y and Λ_x are $n \times q$ and $k \times q$ orthonormal matrices respectively, and D is a $q \times q$ diagonal matrix. The model thus assumes that the codependency between y_t and x_t is explained by a total of $q \leq \min\{n, k\}$ latent factors, whose estimation is the underlying objective. The idiosyncratic component ξ_t captures the rest of the variation in y_t. Occasionally $\Lambda_x' x_t$ or $D\Lambda_x' x_t$ are called PLS factors. In this paper, we regard the $D\Lambda_x' x_t$ as the PLSR estimates of f_t in Equation (1a). Note that moving from x_t to these factors is the dimension reduction step of this approach.

There are multiple algorithms to estimate the PLSR of (3) (see Lohmöller 2013). One possible choice is the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm introduced to PLS by Wold (1975). Since we have only two matrices of data (the response and the explanatory variables), Höskuldsson (1988)'s formulation of NIPALS, given by Algorithm 1, is perhaps the simplest. Let $X = (x_1, \dots, x_T)'$ and $Y = (y_1, \dots, y_T)'$ be the standardized data matrices, and $\Lambda_x = (\lambda_{x1}, \dots, \lambda_{xq})$ and $\Lambda_y = (\lambda_{y1}, \dots, \lambda_{yq})$. The Algorithm 1 finds the projection directions $\lambda_{xr}'\lambda_{xr} = 1$ and $\lambda_{yr}'\lambda_{yr} = 1$ such that $\mathrm{Cov}(X\lambda_{xr}, Y\lambda_{yr})^2$ is maximized (Höskuldsson 1988).[2]

ALGORITHM 1 | Höskuldsson (1988) NIPALS algorithm.

Input: $X \in \mathbb{R}^{T\times k}$, $Y \in \mathbb{R}^{T\times n}$

Output: $\Lambda_x \in \mathbb{R}^{k\times q}$, $\Lambda_y \in \mathbb{R}^{n\times q}$, $D \in \mathbb{R}^{q,q}$

Require: $q > 0$

1: $D \leftarrow \mathbf{0}_{q,q}$
2: **for** $r = 1$ to q **do**
3: $u_r \leftarrow$ the first column of Y.
4: **while** u_r has not converged **do**
5: $\lambda_{xr} \leftarrow X'u_r$ and scale λ_{xr} to unit length.
6: $v_r \leftarrow X\lambda_{xr}$.
7: $\lambda_{yr} \leftarrow Y'v_r$ and scale λ_{yr} to unit length.
8: $u_r \leftarrow Y\lambda_{yr}$.
9: **end while**
10: $p_r \leftarrow X'v_r/(v_r'v_r)$ and $d_r \leftarrow u_r'v_r/(v_r'v_r)$.
11: $X \leftarrow X - v_r p_r'$ and $Y \leftarrow Y - d_r v_r \lambda_{yr}'$.
12: $[\Lambda_y]_{\cdot,r} \leftarrow \lambda_{yr}$, $[\Lambda_{xr}]_{\cdot,r} \leftarrow \lambda_x$ and $[D]_{r,r} \leftarrow d_r$.
13: **end for**

Verbally Algorithm 1 works by estimating factors based on Y (denoted by u_r, row 8) and factors based on X (denoted by v_r, row 6) in an alternating manner. The loadings of u_r are estimated by regression of Y on v_r (row 7) and the loadings of v_r are estimated by regression of X on u_r (row 5). This is repeated until convergence. This movement between X and Y allows us to skip over factors that only affect X or Y but not both. Subsequent pairs of factors beyond the first can be estimated by rerunning the algorithm for the residuals of Y and X (row 11). The diagonal matrix D of Equation (3) connects u_r and v_r. It is estimated by univariate regressions between the factors (rows 10 and 12).

Note that the factor loadings Λ_y and Λ_x are standardized to be orthonormal. Meanwhile, the factors are not standardized. This means that the covariance between factors is directly related to how much of the variance of y_t is explained by factors estimated based on x_t. In the Algorithm 1, this covariance is $\frac{u_r'v_r}{T}$. Since Y is standardized, the average variance is 1. This means that the share of covariance extracted by the factor r relative to the variance of Y is

$$\phi_r = \frac{|u_r'v_r|}{Tn}. \tag{4}$$

A scree plot of ϕ_r for each factor could then be used to choose the number of factors. Examples of these plots can be found in Section 4.4. Alternatively, we can stop estimating more factors once enough of the covariance is extracted, that is, the cumulative sum of (4) is enough (say 25%).

When y_t is a scalar (i.e., $n = 1$), NIPALS estimates the best linear predictor $\Lambda_x' x_t$ of y_t when q is taken to be the number of

nonzero eigenvalues of Cov(X) such that the corresponding eigenvectors are not orthogonal to the covariance between X and Y (Helland 1990). Under PCR, the latter property is not present, which means that the PCR typically requires the estimation of more factors.

PLS can be used to estimate a non-dynamic factor model: Assume that x_t is generated by factors f_t and that y_t is generated by a subset of these factors (and possibly some additional factors that are orthogonal to f_t) like in (1a). Also, assume that the idiosyncratic components of y_t do not correlate with the idiosyncratic component of x_t or the subset of factors that generate y_t. Lastly, assume that the idiosyncratic components are homogeneous. Under these assumptions, PLS can consistently estimate the factor model (Helland 1990). The same variance or homogeneity assumption is, in general not guaranteed to hold. One common partial solution is to standardize the time series to have a variance of 1. Additionally, when dealing with time series, the idiosyncratic components of x_t and y_t are likely to be correlated. However, simplifying assumptions regarding the correlation structures of the idiosyncratic components are essential features of the ML and especially the QML method for the estimation of DFMs (see, for example, Doz et al. 2012).

The PLSR has been used in economic forecasting, although it still remains a marginal method when compared to the DFM. A notable technical paper by Groen and Kapetanios (2016) shows that when the data generating process is a factor model and a set of general assumptions are made, asymptotically the PLSR and the PCR are equivalent. However, (Groen and Kapetanios 2016) argue that in finite datasets, the PLSR is more parsimonious and that it is optimal even when the factors are particularly weak or when irrelevant data is added to the set of predictors. They also apply their method to macroeconomic time series. Furthermore, Kelly and Pruitt (2013, 2015) develop a three-pass-regression-filter (3PRF), which differs from PLSR mainly in its way of dealing with constant terms and variances of the predictors. They apply the method to forecasting both macroeconomic and financial data with great success. Fuentes et al. (2015) develop a sparse version of the PLSR and use it for forecasting macroeconomic data, and Eickmeier and Ng (2011) forecast economic activity in New Zealand with the PLSR. In summary, most of the previous studies have focused on forecasting a single variable, usually with monthly data, and ignore matters of real-time data availability.

The closest paper to ours is Hepenstrick and Marcellino (2019), who also use the PLSR, or more specifically the 3PRF, to nowcast macroeconomic variables using mixed-frequency data. However, their focus is on modeling a single endogenous variable at a time. To deal with mixed frequency target, they use lower-frequency regressions, when necessary, which is what we also do. To actually forecast or nowcast the target, they use U-MIDAS, which is a time series regression method for a mixed frequency target. They deal with lower frequency, or in general missing predictors, by using an alternative method to nowcast them, which is necessary since their method allows only for a single endogenous variable.

3 | High-Dimensional Time Series Modeling With PLSR

3.1 | Vector Autoregressive PLSR (VAR-PLSR)

We next extend model (3) to a vector autoregressive PLSR (VAR-PLSR) model with centered time series $y_t \in \mathbb{R}^n$ that have been scaled to have unit variance. Let us use two lags of y_t as predictors for notational and illustrative purposes. Since y_t is a vector, and it is a response and its 2 lags are predictors, this is called VAR(2). More or fewer lags can readily be considered. Formally, $x_t' = (y_{t-1}', y_{t-2}')$:

$$y_t = \Lambda_y D \Lambda_x' \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \xi_t, \ t = 1, \ \dots, T, \tag{5}$$

where Λ_y is a $n \times q$ matrix, Λ_x is a $2n \times q$ matrix, and D is a $q \times q$ diagonal matrix. As noted above, $f_t = D\Lambda_x'(y_{t-1}', y_{t-2}')'$ can be seen as PLS factors like in (3).

Mixed frequency data requires some special consideration. However, expressing the model as a factor model allows us to use the conventional method introduced by Mariano and Murasawa (2003). Consider decomposing (5) into two equations:

$$y_t = \Lambda_y f_t + \xi_t, \tag{6a}$$

$$f_t = D\Lambda_x' \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} = D\Lambda_x' \begin{bmatrix} \Lambda_y f_{t-1} \\ \Lambda_y f_{t-2} \end{bmatrix} + D\Lambda_x' \begin{bmatrix} \xi_{t-1} \\ \xi_{t-2} \end{bmatrix}, \ t = 1, \ \dots, T. \tag{6b}$$

Now we can consider a mix of monthly and quarterly variables, where the quarterly variables are differences of trending variables (such as GDP). Let $y_t^{(Q)}$ be the part of y_t that includes quarterly observed variables, $\Lambda_y^{(Q)}$ the rows of Λ_y that correspond to these variables, and, finally, $\xi_t^{(Q)}$ the idiosyncratic components of these variables. The remaining y_t, Λ_y and ξ_t correspond to monthly time series. Equation (6a) for these quarterly variables can be written as

$$y_t^{(Q)} = \Lambda_y^{(Q)}(f_t + 2f_{t-1} + 3f_{t-2} + 2f_{t-3} + f_{t-4})/3 + \xi_t^{(Q)}, \ t = 1, \ \dots, T. \tag{7}$$

If the quarterly variables are not trending, a simple three lag filter can be used.

While Equation (6a) is easy to adjust for $y_t^{(Q)}$, this not the case for (6b). Hepenstrick and Marcellino (2019) recommend interpolating lower frequency observations. However, the easiest solution is to simply omit $y_t^{(Q)}$ in (6b). This would make Λ_x sparse, i.e., the rows corresponding to the quarterly series would be set to zero. We follow the latter approach for simplicity and this is also the approach Hepenstrick and Marcellino (2019) take in their forecasting exercises. Additionally, if we assume that the data generating process is a DFM, this loss of information has a negligible effect as long as the number of higher frequency predictors is large enough, because of the asymptotic properties of PLS (Groen and Kapetanios 2016).

We write the conditional expectation of y_{t+h} at time t, given the information available at that date, as $y_{t+h|t}$. Forecasting is done by alternating between forecasting f_{t+h} using (6b) and forecasting y_{t+h} using (6a):

$$f_{t+h|t} = D\Lambda_x'\left(y_{t+h-1|t}', y_{t+h-2|t}'\right)',$$
$$y_{t+h|t} = \Lambda_y f_{t+h|t} + \xi_{t+h|t}.$$

If the idiosyncratic components are not autocorrelated then $\xi_{t+h|t} = 0$ when $h > 0$. Because the estimation of such autocorrelation structure would be difficult without maximum likelihood estimation, we will assume that $\xi_{t+h|t} = 0$. Additionally, if the idiosyncratic components are not autocorrelated, forecasts at any horizon can be constructed easily with the companion form representation:

$$\begin{bmatrix} y_{t+h|t} \\ y_{t+h-1|t} \end{bmatrix} = \begin{bmatrix} \Lambda_y D\Lambda_x' & \\ I_n & 0_n \end{bmatrix}^h \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}, \tag{8}$$

where I_n is an $n \times n$ identity matrix and 0_n is an $n \times n$ zero matrix. When dealing with mixed frequency data, the recursion in (7) must be used.

To estimate VAR-PLSR when there are no missing observations and all observations have the same data frequency, Algorithm 1 can be used as is by defining the data matrices as $Y = (y_{p+1}', \dots, y_t')'$ and $X = ((y_1', \dots, y_p')', \dots, (y_{T-p}', \dots, y_{T-1}')')'$. However, when some observations are missing or some observations are quarterly and some monthly, a few modifications are necessary. We recommend Algorithm 2, which is a simple variation of the original algorithm by Höskuldsson (1988) that uses lower frequency regressions and applies the methods of Mariano and Murasawa (2003) when applicable. Let $\underline{y}_i = (y_{p+1,i}, \dots, y_{T,i})'$ and $\tilde{\underline{y}}_i$ is the same vector with missing observations omitted. Also let $\tilde{v}_{ri}$ be v_r with those components removed that correspond to the missing observations in $\underline{y}_i$. The vectors $\tilde{\underline{x}}_j$ and $\tilde{u}_{rj}$ are obtained analogously from X. Finally, let $\widehat{Y}$ and $\widehat{X}$ denote Y and X with missing observations replaced by zeros. Additionally, when dealing with mixed frequency data, let $\tilde{\underline{y}}_i$ not include the first four months if the i th series is quarterly. Likewise, let $\tilde{v}_{rit}^{(Q)} = (v_{rt} + 2v_{rt-1} + 3v_{rt-2} + 2v_{rt-3} + v_{rt-4})/3$ for all t for when the quarterly series is observed. Note that with mixed frequency data, the quarterly series are omitted from X.

In plain language, Algorithm 2 works by alternating between X and Y, just like Algorithm 1. However, the algorithm now iterates through the individual series (rows 4–10 and 13–19) to account for mixed frequencies. For Y, the loadings are estimated just like in Algorithm 1 when the series is monthly (row 15), but when the data is quarterly, the factors are aggregated just like in Mariano and Murasawa (2003) (row 17). For X the loadings of monthly series are treated the same (row 6), but the quarterly series are omitted (row 8). The factors themselves can be estimated just like in Algorithm 1 (rows 12 and 21).

It is typical that all series used in PLS are standardized. However, if there are any quarterly series, these are best scaled to have a variance of 19/9 so that they match the variance of $\tilde{v}_{rit}^{(Q)}$. Otherwise, the loadings for the quarterly series are likely to be very small. Of course, when the nowcasts and forecasts are rescaled, this needs to be taken into account. This aggregation means that Equation (4) is not necessarily accurate in describing the share of covariance extracted by a factor relative to total variance of data, but it is still instructive especially if the share of quarterly variables is small.

ALGORITHM 2 | NIPALS algorithm for mixed frequency time series data.

Input: $X \in \mathbb{R}^{(T-p)\times(pn)}$, $Y \in \mathbb{R}^{(T-p)\times n}$

Output: $\Lambda_x \in \mathbb{R}^{(pn)\times q}$, $\Lambda_y \in \mathbb{R}^{n\times q}$, $D \in \mathbb{R}^{q\times q}$

Require: $q > 0$

1: **for** $r = 1$ to q **do**
2: $u_r \leftarrow$ the first column of $\widehat{Y}$.
3: **while** u_r has not converged **do**
4: **for** $j = 1$ to pN **do**
5: **if** j:th series in monthly **then**
6: $\lambda_{xrj} \leftarrow \tilde{\underline{x}}_j' \tilde{u}_{rj}$.
7: **else**
8: $\lambda_{xrj} \leftarrow 0$.
9: **end if**
10: **end for**
11: Scale λ_{xr} to unit length.
12: $v_r \leftarrow \widehat{X}\lambda_{xr}$.
13: **for** $i = 1$ to N
14: **if** i:th series is monthly **then**
15: $\lambda_{yri} \leftarrow \tilde{\underline{y}}_i' \tilde{v}_{ri}$
16: **else**
17: $\lambda_{yri} \leftarrow \tilde{\underline{y}}_i' \tilde{v}_{ri}^{(Q)}$
18: **end if**
19: **end for**
20: Scale λ_{yr} to unit length
21: $u_r \leftarrow \widehat{Y}\lambda_{yr}$.
22: **end while**
23: $p_r \leftarrow \widehat{X}' v_r/(v_r'v_r)$ and $d \leftarrow u_r'v_r/(v_r'v_r)$.
24: $\widehat{X} \leftarrow \widehat{X} - v_r p_r'$ and $\widehat{Y} \leftarrow \widehat{Y} - d_r v_r \lambda_{yr}'$.
25: $[\Lambda_y]_{\cdot,r} \leftarrow \lambda_{yr}$, $[\Lambda_x]_{\cdot,r} \leftarrow \lambda_{xr}$ and $[D]_{r,r} \leftarrow d_r$.
26: **end for**

VAR-PLSR solves some of the challenges that PLSR has when it is applied to time series. Most previous studies, such as Eickmeier and Ng (2011), Kelly and Pruitt (2015) and Groen and Kapetanios (2016), ignore publication lags prevalent in macroeconomic data but since in VAR-PLSR all predictors are endogenous, missing data can be easily filled in. They also convert

all the data into the same frequency. While we do omit lower frequency series from predictors, they are used in the estimation of the model. Hepenstrick and Marcellino (2019) solve the mixed frequency problem, but they have to resort to ad hoc methods to deal with the ragged edge. Their method also has to forecast variables directly, while VAR-PLSR can form recursive forecasts easily using Equation (8). Also, while we now ignore this possibility, VAR-PLSR could easily be used for structural analysis like any other structural vector autoregression (SVAR).

3.2 | Estimating DFM With PLS (PLS-DFM)

We next show how the decomposed PLSR of Equations (6a) and (6b) can be turned into an actual DFM with a couple of alterations. Firstly, take $\Lambda_x' = (\Lambda_{1x}', \Lambda_{2x}')$, where Λ_{1x} and Λ_{2x} are $n \times q$ matrices. Then we can write Equation (6b) as

$$f_t = D\Lambda_{1x}'\Lambda_y f_{t-1} + D\Lambda_{2x}'\Lambda_y f_{t-2} + D\Lambda_{1x}'\xi_{t-1} + D\Lambda_{2x}'\xi_{t-2}.$$

In an actual DFM, the idiosyncratic component has no effect on the factors. Thus, we substitute $D\Lambda_{1x}'\xi_{t-1} + D\Lambda_{2x}'\xi_{t-2}$ with innovations w_t that are independent of ξ_t. Strictly speaking, this is not true, but it is a reasonable assumption when n is large as long as the loadings $D\Lambda_x'$ and the variance in ξ_t is not concentrated on just a few components.[3] This is obvious if the components of ξ_t and ξ_{t-1} are uncorrelated and have same variances, but it is also reasonable as long as they do not have a factor structure that is non-orthogonal to f_t and the variances are positive but finite. This results in a DFM:

$$y_t = \Lambda_y f_t + \xi_t \tag{9a}$$

$$f_t = D\Lambda_{1x}'\Lambda_y f_{t-1} + D\Lambda_{2x}'\Lambda_y f_{t-2} + w_t. \tag{9b}$$

We call a DFM estimated with PLS a PLS-DFM.

The current setup allows us to estimate the loadings Λ_y and $\Lambda_x D$ as well as the transition matrices $D\Lambda_{1x}'\Lambda_y$ and $D\Lambda_{2x}'\Lambda_y$, using PLS while still allowing us to use various tools that are available to common DFMs, such as Kalman filter. In particular, we want to address the problem of missing data (such as the ragged edge). Additionally, there may still be unestimated and unmodeled factors that affect ξ_t, but since they have no predictive power, they are omitted. This omission allows for more parsimonious models.

The concrete implementation is as follows:

1. Just like in VAR-PLSR, use NIPALS of Algorithm 1 (Algorithm 2 if dealing with mixed frequency or missing data) to estimate Λ_y, D, and Λ_x.
2. Use the residuals ξ_t in VAR-PLSR to estimate $\text{Cov}(\xi_t)$. Set off-diagonals to zero. Use the residuals from Equation (6b) to estimate $\text{Cov}(w_t)$.
3. The factors f_t are estimated with Kalman filter using the parameter estimates of step 1 (see Equations 9a and 9b).

The above construction has similarities to the two-step estimation procedure of Doz et al. (2011). The VAR-PLSR is the analogous first step that estimates the loadings and the transition matrix. However, unlike in Doz et al. (2011), PLS also gives the transition matrix immediately without a need for a separate VAR. The covariance matrices of ξ_t and w_t can be estimated from the residuals of VAR-PLSR (step 2 above), much like in Doz et al. (2011). The last step is the Kalman filter.

Just like in (7) the mixed data frequency must be accounted for Equation (9a):

$$\begin{aligned} y_t^{(Q)} = \Lambda_y^{(Q)} &(f_t + 2f_{t-1} + 3f_{t-2} + 2f_{t-3} + f_{t-4})/3 \\ &+ (\xi_t^{(Q)} + 2\xi_{t-1}^{(Q)} + 3\xi_{t-2}^{(Q)} + 2\xi_{t-3}^{(Q)} + \xi_{t-4}^{(Q)})/3. \end{aligned}$$

Because of this, it is more practical to use a state vector

$$s_t = \left(f_t', f_{t-1}', f_{t-2}', f_{t-3}', f_{t-4}', \xi_t', \xi_{t-1}^{(Q)\prime}, \xi_{t-2}^{(Q)\prime}, \xi_{t-3}^{(Q)\prime}, \xi_{t-4}^{(Q)\prime}\right)'$$

in the Kalman filter, instead of using the factors f_t directly. For more details about mixed frequency DFMs, see Mariano and Murasawa (2003) and Bańbura and Modugno (2014).

Turning VAR-PLSR into a DFM also has advantages over other PLS methods. As mentioned in the previous sections, most PLSR literature has ignored mixed data frequency and publication lags. Turning VAR-PLSR into a DFM allows us to use Kalman filter to address these issues. Hepenstrick and Marcellino (2019) recommend a number of ad hoc solutions to the publication lag issue. The best method they try is to use Kalman filter, where factor loadings are the coefficient estimates from PLSR and the transition matrix is taken from fitting a VAR to their factor estimates. In other words, they end up using a DFM even though their model is not estimated to work as a DFM like our approach is.

3.3 | Comparison of Methods

The benefits of using VAR-PLSR or PLS-DFM instead of a DFM estimated with either principal components (2s-DFM) or ML is that the factors and their loadings are specifically estimated to predict y_t rather than trying to just capture cross-sectional comovements of y_t. If there are strong comovements in the idiosyncratic component or non-persistent factors that we are not interested in, this means that we can have a smaller q and a more parsimonious model. As noted by Groen and Kapetanios (2016), this property of PLS can be particularly valuable in small samples.

Parsimony also makes the model specification easier: If we suspect that the correct number of factors is between, say, 1 and 6, but only half of these are persistent and thus useful for forecasting, it is easier to choose the number of persistent factors, which is between 1 and 3, than it is to choose the number of overall factors, which is between 1 and 6. Furthermore, if T is very small and n very large, it may become impossible to estimate all factors. Indeed, PLS can be seen as an alternative approach to selecting just a subgroup of predictors through prescreening (see Boivin and Ng 2006).

In contrast to 2S-DFM, the NIPALS algorithm described can account for mixed frequency data. The NIPALS also estimates the β coefficient matrices of Equation (1b) in the same step as Λ, which is computationally attractive. When compared to ML, the PLS based

methods are much more practically feasible, as the EM-algorithm needed for ML is notoriously slow (see, for example, Ng et al. 2011).

When comparing the two methods of Sections 3.1 and 3.2, PLS-DFM can generally be expected to perform better when the data generating process is a DFM, since Kalman filter finds the mean squared error optimal factor estimates. However, under some exotic data generating processes, VAR-PLSR may perform better. Additionally, as Equation (6b) shows, in VAR-PLSR idiosyncratic components can influence factors and other variables. This can be beneficial in small open economies like Finland, which is investigated in Section 5: These economies (i.e., their factors) are strongly influenced by global dynamics that are often only represented with just a few indicators in the data sets. It can thus be useful if idiosyncratic components of these variables could influence the factors directly. This is impossible in typical DFMs but is possible in VARs that have inbuilt factors, like VAR-PLSR.

Lastly, VARs are more popular than DFMs, and they can be often easier to interpret. VAR-PLSR offers an approach to econometric modeling that works like a VAR while mimicking a DFM. This allows an econometrician to estimate impulse responses, forecast error variance decompositions and other interesting statistics more easily.

4 | Simulation Results

Before examining empirical results in a real-world forecasting situation in Section 5, we consider several simulation experiments when comparing PLSR-based approaches to existing methods.

4.1 | Forecasting Experiment With DFM Generated Data

We replicate the simulation experiment of Doz et al. (2012), which is also examined in Bańbura and Modugno (2014), now containing our PLS-based approach. We include the methods described in the previous section, as well as the conventional DFM approach via ML described in Bańbura and Modugno (2014).

The data generating process (DGP) is the following approximate dynamic factor model:

$$\begin{aligned} y_t &= \Lambda f_t + \xi_t, \qquad t = 1, \ldots, T, \\ f_t &= \beta f_{t-1} + w_t, \; w_t \sim nid(0, I_q), \\ \xi_t &= \Gamma \xi_{t-1} + \nu_t, \qquad \nu_t \sim nid(0, \Sigma), \end{aligned}$$

where

$$\begin{aligned} \Lambda_{ij} &\sim nid(0,1), \; i = 1, \ldots, n, \; j = 1, \ldots, q, \\ \beta &= I_q \rho, \\ \Gamma &= I_n \alpha, \\ \Sigma_{ij} &= \tau^{|i-j|}(1-\alpha^2)\sqrt{\gamma_i \gamma_j}, \\ \gamma_i &= \frac{b_i}{1-b_i} \frac{1}{1-\rho^2} \sum_{j=1}^{q} \Lambda_{ij}, \\ b_i &\sim U([u, 1-u]). \end{aligned}$$

Here, τ determines the cross-correlation and b_i the heterogeneity of the idiosyncratic component. More specifically, parameter b_i determines the variance of the idiosyncratic shock relative to the total variance of y_i. This is sampled from a uniform distribution with a mean of 0.5, and thus, the idiosyncratic component determines on average half of the total variance. The variance of the idiosyncratic component i itself is γ_i. If $\tau = 0$, the model is an exact DFM, and if $u = 0.5$, $b_i = 0.5$ and the idiosyncratic component is homogeneous. If τ is increased, the idiosyncratic component becomes cross-correlated, but since Σ is a Toeplitz matrix, this will not imply a factor structure. Furthermore, increasing α increases the autocorrelation of the idiosyncratic component and increasing ρ increases the persistence of the factors. Choosing $\alpha = 0$ leads to white noise idiosyncratic components. We considered different combinations of ρ, α, τ, in conjunction with various numbers q of true factors and numbers $\hat{q}$ of estimated factors. However, we keep $u = 0.1$ as this corresponds best to reality, where idiosyncratic components are generally not homogeneous. This is also the setting used by Bańbura and Modugno (2014).

Table 1 reports the root mean squared forecast errors (RMSFE) of the forecast simulations relative to the RMSFEs of a naive forecast using the sample mean as the forecast. These are constructed by taking the one-period forecast errors of all n series, squaring them, and then taking the average over all n series and 1000 replicates of the simulation before taking the square root. The final root mean square errors of different methods are divided by those of the sample mean. The methods considered are the vector autoregressive PLSR (VAR-PLSR, defined in Section 3.1), the DFM estimated by PLS (PLS-DFM, defined in Section 3.2), the DFM estimated with the two-step estimator (2s-DFM) (see Doz et al. 2011) and the DFM estimated by ML as in Bańbura and Modugno (2014) (ML-DFM). ML-DFM allows for autocorrelated idiosyncratic components if $\alpha \neq 0$ in the DGP, whereas the 2s-DFM, VAR-PLSR and PLS-DFM do not.

The following list summarizes the implementation procedure of the 2s-DFM and ML-DFM methods:

- 2S-DFM (two-step algorithm, Doz et al. 2011):
  1. Estimate $\hat{q}$ principal components and eigenvectors of the sample covariance matrix of y_t. Eigenvectors are used as Λ, β is estimated with a VAR(1) on the principal components, and residuals of PCA and VAR(1) are used to estimate $\text{Cov}(\xi_t)$ and $\text{Cov}(w_t)$. Off-diagonal terms of $\text{Cov}(\xi_t)$ are set to zero.
  2. The factors f_t are estimated with Kalman filter using the parameter estimates of step 1.
- ML-DFM (EM-algorithm, see for example Bańbura and Modugno 2014):
  1. The algorithm is initialized with the two-step algorithm.
  2. (M-step): Assume f_t to be fixed and estimate Λ, β, $\text{Cov}(\xi_t)$ and $\text{Cov}(w_t)$ with maximum likelihood.
  3. (E-step): Estimate factors f_t with Kalman filter using the parameter estimates the M-step.
  4. Repeat steps 2 and 3 until convergence.

Table 1 includes four panels. Each panel corresponds to a different data generating DFM: Panel A is a one factor exact DFM with a highly persistent factor. Panels B, C and D are all larger

TABLE 1 | Forecast simulation results: Only persistent factors.

Panel A: $\rho = 0.9, \alpha = 0, \tau = 0, u = 0.1, q = \widehat{q} = 1$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.732	0.741	0.743	0.655	0.658	0.658	0.741	0.743	0.744	0.667	0.662	0.662
100	0.706	0.716	0.724	0.629	0.629	0.631	0.718	0.717	0.725	0.639	0.631	0.633

Panel B: $\rho = 0.9, \alpha = 0, \tau = 0, u = 0.1, q = \widehat{q} = 3$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.777	0.766	0.764	0.697	0.683	0.680	0.771	0.756	0.755	0.725	0.694	0.689
100	0.750	0.732	0.734	0.676	0.652	0.651	0.756	0.734	0.732	0.699	0.660	0.656

Panel C: $\rho = 0.7, \alpha = 0, \tau = 0, u = 0.1, q = \widehat{q} = 3$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.966	0.962	0.961	0.903	0.889	0.886	0.963	0.959	0.957	0.922	0.900	0.895
100	0.957	0.950	0.952	0.892	0.873	0.873	0.960	0.952	0.952	0.911	0.884	0.880

Panel D: $\rho = 0.9, \alpha = 0.5, \tau = 0.5, u = 0.1, q = \widehat{q} = 3$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.765	0.756	0.754	0.634	0.626	0.625	0.751	0.746	0.744	0.685	0.682	0.677
100	0.737	0.729	0.729	0.613	0.599	0.598	0.742	0.729	0.728	0.663	0.653	0.651

Note: The entries in Panels A–D are simulated root mean squared forecast errors (RMSFE) of 2s-DFM, ML-DFM, VAR-PLSR, and PLS-DFM divided by those of just using the sample mean as the prediction. Simulation is repeated 5000 times (without missing data). Panels correspond to different data generating DFMs, n is the size of the cross-section and T is the length of the training sample. The number of estimated factors is $\widehat{q}$. Whenever the DGP has autocorrelated idiosyncratic components ($\alpha \neq 0$), so does the ML-DFM. The forecast horizon in these simulation experiments is throughout one period.

three factor DFMs. Panel B is an exact factor model with three highly persistent factors, Panel C has less persistent factors and Panel D is an approximate model with cross- and autocorrelation present in the idiosyncratic components. It is important to note that the Panels A-C match exactly the approximating ML-DFM model. That is, we can expect that any alternative approach will yield larger prediction errors than this model corresponding the DGP. Therefore, the question of interest is that how large are the forecast losses, especially with the PLS-DFM, compared with the "correct" process that the ML-DFM represents.

Based on Table 1, we can conclude that the ML yields smaller RMSFEs than PLS in forecasting, although the difference gets smaller as T and n grow. The former makes sense as the DGP matches the model of the ML, and the latter makes sense since the approaches are asymptotically equivalent. However, the PLS-DFM outperforms VAR-PLSR, which suggests that using the Kalman filter to estimate the factors is indeed beneficial. Also, the PLS-DFM outperforms the two-step estimator.

In Appendix A, we report the differences in average computation times required for different estimation methods. On average the computation times of PLS-DFM are about q times those of 2s-DFM whereas ML-DFM are on average about 10 times those of 2s-DFM. Meanwhile, VAR-PLSR is always quicker than PLS-DFM. Overall, the PLS-DFM seems to provide good compromise between the computational burden of the method and forecasting performance.

Appendix B reports the results for the same DGPs as Table 1 but with a forecast horizon $h = 3$. The RMSFEs are higher at $h = 3$, but the relative errors between the four methods are generally the same as in Table 1.

4.2 | Mixed Frequency Forecasting Simulation

A common challenge posed by macroeconomic time series is the mixed frequency of variables, most often quarterly (such

as GDP) and monthly (such as CPI). The Algorithm 2 provides a method of estimating a DFM or VAR with PLS when some of the data is monthly and the rest quarterly. In this section, we analyze the forecast performance of VAR-PLSR and PLS-DFM when 20% of the data is quarterly and the rest is monthly.

Table 2 reports RMSFEs for ML-DFM, VAR-PLSR and PLS-DFM when the data is generated by the same DGPs as in Table 1, except that 20% of variables are quarterly. The 2s-DFM is omitted as it cannot incorporate mixed frequency data. Table 2 only reports RMSFEs for the quarterly frequency data.

The results shown in Table 2 suggest that PLS-based methods tend to often do better than ML-DFM in the presence of mixed frequency data. However, the asymptotics as T and n grow seem less clear. A possible reason may be the omission of the lower frequency variables among the predictors in Algorithm 2.

4.3 | Monte Carlo Experiment With a Mix of Persistent and Non-Persistent Factors

The key feature of PLS is that it can ignore factors that do not help in prediction. To study this, let us fix the number of factors at 3 and consider the following changes to the DGP in Section 4.1:

$$\beta=\begin{bmatrix}\rho & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix},\ w\sim N(0,\Xi) \text{ and } \Sigma_{ij}=\tau^{|i-j|}\left(1-\left(\frac{\alpha}{3}\right)^2\right)\sqrt{\gamma_i\gamma_j},$$

where

$$\Xi=\begin{bmatrix}1 & 0 & 0\\ 0 & \frac{1}{1-\rho^2} & 0\\ 0 & 0 & \frac{1}{1-\rho^2}\end{bmatrix}.$$

TABLE 2 | Forecast simulation results: Mixed frequency.

Panel Am: $\rho=0.9,\alpha=0,\tau=0,u=0.1,q=\widehat{q}=1$

	ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n		
T	10	50	100	10	50	100	10	50	100
51	0.915	0.916	0.919	0.885	0.865	0.88	0.643	0.722	0.756
99	0.926	0.923	0.923	0.862	0.912	0.92	0.626	0.757	0.785

Panel Bm: $\rho=0.9,\alpha=0,\tau=0,u=0.1,q=\widehat{q}=3$

	ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n		
T	10	50	100	10	50	100	10	50	100
51	0.913	0.911	0.910	0.793	0.845	0.853	0.625	0.657	0.676
99	0.936	0.926	0.924	0.844	0.897	0.901	0.628	0.680	0.713

Panel Cm: $\rho=0.7,\alpha=0,\tau=0,u=0.1,q=\widehat{q}=3$

	ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n		
T	10	50	100	10	50	100	10	50	100
51	0.948	0.944	0.943	0.947	0.974	0.967	0.814	0.794	0.797
99	0.966	0.958	0.956	0.940	0.959	0.962	0.805	0.782	0.802

Panel Dm: $\rho=0.9,\alpha=0.5,\tau=0.5,u=0.1,q=\widehat{q}=3$

	ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n		
T	10	50	100	10	50	100	10	50	100
51	0.825	0.771	0.770	0.881	0.897	0.902	0.678	0.726	0.743
99	0.892	0.843	0.836	0.921	0.946	0.950	0.680	0.739	0.772

Note: In the table (Panels A2–D2), 20% of the variables have a quarterly frequency and 80% monthly. The RMSFEs are reported for the quarterly variables with horizon of $h=3$. See Table 1 for details.

What these changes imply is that now there is one persistent and two non-persistent factors. Additionally, all factors have the same unconditional variance.

Table 3 reports the same summary statistics (RMSFEs) as Table 1 for three simulations using the above alterations to the DGP, where there are three factors in total in the DGP, one of which is persistent and estimated in practice ($\hat{q} = 1$). In Panel E1 data is generated by an exact DFM with one highly persistent factor. In Panel F1 the persistent factor is less persistent, and in Panel G1 the GDP is an approximated DFM with autocorrelated idiosyncratic components. The results suggest that now PLS-DFM outperforms the other approaches even when the DGP is an approximate DFM. The fact that it outperforms the VAR-PLSR suggests that using the Kalman filter is beneficial. Furthermore, the VAR-PLSR outperforms 2s-DFM in Panels E1 and G1 and ML-DFM in Panel E1. Overall, in these simulations, PLS seems to be superior over the existing approaches.

One possible explanation for why the PLS based methods perform so well in Table 3 is that they estimate the one persistent factor better than the alternatives. To examine this possibility, we investigate if the PLS estimates (finds) the persistent factors better than the alternative approaches. To do this, we use a popular trace-R^2 statistic for the persistent factors (Stock and Watson 2002; Doz et al. 2012; Bańbura and Modugno 2014). The formula[4] is

$$R^2_{\text{trace}} = \frac{\text{Trace}(F_T' \widehat{F}_{T|T} (\widehat{F}_{T|T}' \widehat{F}_{T|T})^{-1} \widehat{F}_{T|T}' F_T)}{\text{Trace}(F_T' F_T)},$$

where

$$F_T = (f_{1,1}, \ldots, f_{T,1})',$$

and $f_{t,1}$ is the first factor, that is, the persistent factor, at time t. The interpretation of this statistic is the same as for a conventional R^2: Higher R^2_{trace} means that the movements of the real factor are captured by the estimated factor. Moreover, for the 2s-DFM, ML-DFM and the PLS-DFM, $\widehat{F}_{T|T}$ are estimated using the Kalman smoother, so that the whole data up to T is used for estimation. Meanwhile, the estimation of VAR-PLSR also uses information up to T although the factor estimates for any period t depend on information observed after t only through the estimation of parameters in Equation (5).

Table 4 reports the R^2_{trace} statistics for the same three DGPs as considered in Table 3: Panel E2 in Table 4 corresponds to the Panel E1 in Table 3, and so forth. Here, we find that the PLS-DFM delivers R^2_{trace} statistics that are often multiple times those of the 2s-DFM or the ML-DFM. Furthermore, comparing Panels E2 and G2, we see that the R^2_{trace} statistics of the ML-DFM decrease a lot when the DGP switches from an exact one to an approximate one, whereas those of the PLS-DFM are almost unaffected. This suggests that PLS is more robust than the alternatives. Finally, the VAR-PLSR also outperforms the 2s-DFM in all panels and in addition it outperforms the ML-DFM in Panels E2 and G2. However, it is outperformed by the PLS-DFM, which underlines the importance of the Kalman filter. In summary, PLS seems to find the persistent factor better than the competing approaches.

TABLE 3 | Forecast simulation results: Mix of persistent and non-persistent factors.

Panel E1: $\rho = 0.9, \alpha = 0, \tau = 0, u = 0.1, q = 3, \hat{q} = 1$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.980	0.981	0.981	0.969	0.979	0.978	0.933	0.927	0.926	0.919	0.912	0.909
100	0.978	0.982	0.988	0.955	0.968	0.976	0.919	0.911	0.910	0.899	0.885	0.884

Panel F1: $\rho = 0.7, \alpha = 0, \tau = 0, u = 0.1, q = 3, \hat{q} = 1$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.995	0.994	0.991	0.984	0.989	0.987	0.990	0.986	0.985	0.971	0.965	0.962
100	0.993	0.997	0.998	0.983	0.988	0.988	0.987	0.984	0.983	0.969	0.956	0.954

Panel G1: $\rho = 0.9, \alpha = 0.5, \tau = 0.5, u = 0.1, q = 3, \hat{q} = 1$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.978	0.982	0.980	0.928	0.931	0.930	0.929	0.926	0.925	0.914	0.910	0.908
100	0.979	0.982	0.988	0.916	0.912	0.912	0.918	0.910	0.910	0.896	0.884	0.884

Note: In the table (Panels E1–G1), only the first factor is estimated and only one factor is persistent (non-zero autoregressive coefficients) in the DGP. See Table 1 for details.

TABLE 4 | Factor estimation simulation: Mix of persistent and non-persistent factors.

Panel E2: $\rho = 0.9, \alpha = 0, \tau = 0, u = 0.1, q = 3, \widehat{q} = 1$												
	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.179	0.155	0.149	0.206	0.168	0.157	0.504	0.597	0.613	0.546	0.613	0.623
100	0.226	0.214	0.208	0.310	0.242	0.225	0.683	0.795	0.809	0.737	0.810	0.820
Panel F2: $\rho = 0.7, \alpha = 0, \tau = 0, u = 0.1, q = 3, \widehat{q} = 1$												
	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.279	0.279	0.273	0.301	0.286	0.282	0.564	0.704	0.728	0.586	0.708	0.726
100	0.289	0.297	0.294	0.344	0.313	0.303	0.687	0.852	0.876	0.723	0.859	0.881
Panel G2: $\rho = 0.9, \alpha = 0.5, \tau = 0.5, u = 0.1, q = 3, \widehat{q} = 1$												
	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.180	0.156	0.149	0.182	0.158	0.149	0.513	0.599	0.615	0.551	0.616	0.624
100	0.227	0.214	0.210	0.202	0.166	0.150	0.685	0.795	0.810	0.735	0.810	0.820

Note: The entries are the trace R^2_{trace} statistics for the first (persistent) factor. Starting values, f_0, are omitted. See Tables 1 and 3 for details.

Appendix B reports the results of a mixed frequency simulation based on the DGPs of Tables 3 and 4. The results are similar to those of Tables 2,3, and 4.

4.4 | Choosing the Number of Factors

In Section 2.2, it was suggested that ϕ_r of Equation (4) could be used to choose the right number of factors. The panels in Figure 1 show the average scree plot together with 10%, 25%, 75% and 90% quantiles for the DGPs of panels B, Bm and E with $t = 100$ and $n = 50$. The correct number of persistent factors is three in the first two and one in the last one. Thus, we would expect the "elbow" or the point after which the curve turns flat to be right after 3 for B and Bm. For E, the elbow could be at 1 or 3.

Figure 1a,b shows ϕ_r for the DGPs B and Bm. First, we see that they are very similar, which suggests that mixed data has a fairly small effect on ϕ_r. Second, we see a clear elbow around ϕ_4. This would suggest that a researcher using a scree plot would thus have almost always chosen the correct number of factors, which is 3. However, when there is a mix of persistent and non-persistent factors, like in Figure 1c, the location of the elbow is less clear: The curve flattens more smoothly, although there is a huge gap between ϕ_1 and ϕ_2. Consequently, a researcher would probably either choose 1, 2, or 3. Based on this small experiment, we would recommend going with the smaller number when in doubt. Additionally, the quantiles in every panel of Figure 1 suggest that the variance of ϕ_r tends to be quite small.

5 | Empirical Results: Finnish GDP Growth

5.1 | Nowcasting and Forecasting Results

We apply our methods to forecasting and nowcasting the Finnish real GDP growth. By focusing on a smaller country like Finland, we wish to contribute to the literature that focuses on less studied economies like Eickmeier and Ng (2011). Our dataset consisting of 44 predictors is almost the same as that in Itkonen and Juvonen (2017) and Juvonen and Lindblad (2025) Of these series, nine are quarterly and the rest are monthly series, and the sample period starts in February of 1995 and ends in December 2023. The publication lags of various variables are taken into account, but we ignore data revisions, meaning that, for example, GDP is only observed at a two-month lag, and we use the 2024Q4 vintage. See Appendix C and Juvonen and Lindblad (2025) for more details about the dataset.

The models considered are the PLS-DFM, VAR-PLSR and ML-DFM with the number of lags and factors chosen either $p = q = 1$ or $p = q = 3$ as well as PLS-DFM and VAR-PLSR with $p = 1$ or $p = 3$ and q chosen so that $\sum_{r=1}^{q} \phi_r > 25\%$ (we refer to this as just $\phi_r > 25\%$ in the tables to save space). The ML-DFM allows for autocorrelated idiosyncratic components, which is not the case with PLS-DFM. Also just like in Section 4.2, the 2s-DFM is dropped as it cannot incorporate mixed data frequency or missing observations in a straightforward manner. All forecasting computations use a rolling window of 137 months in updating the estimates. This choice of the number of months corresponds to the number of months from

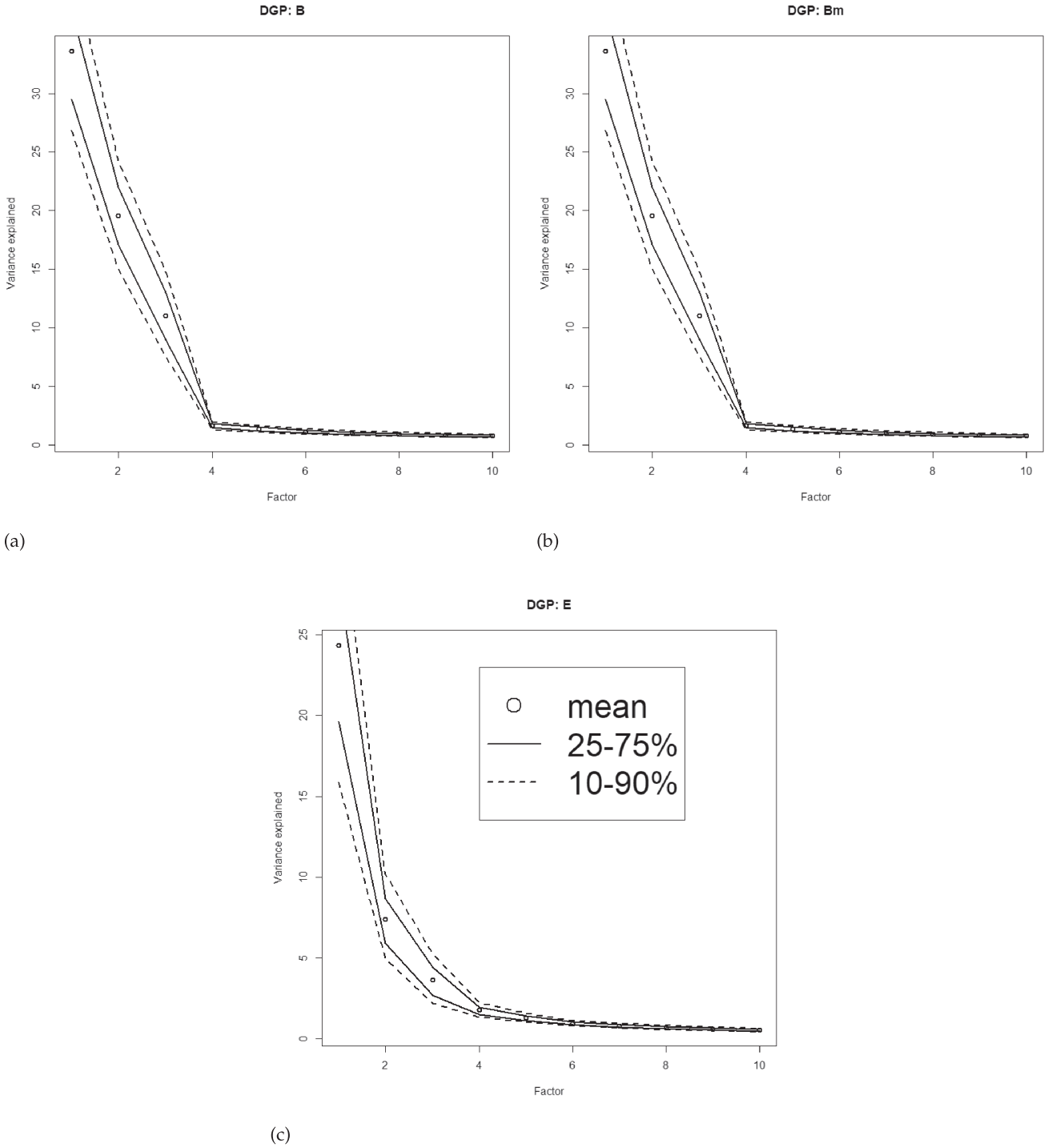

FIGURE 1 | Scree plots show mean and 10%, 25%, 75%, and 90% quantiles for $\phi_r \times 100\%$ of Equation (4) for DGPs of Panels B, Bm, and E with $T = 100$ and $n = 50$. (a) The proposed ϕ_r metric of Equation (4) for DGP of Panel B. Correct number of factors is 3. (b) The proposed ϕ_r metric of Equation (4) for DGP of Panel Bm. Correct number of factors is 3. (c) The proposed ϕ_r metric of Equation (4) for DGP of Panel E. There is one persistent factor and two non-persistent.

February of 1995 to June of 2006, which is the first estimation window in forecast computation.

The VAR-PLSR and PLSR-DFM are also compared to two benchmark models: an autoregressive model using quarterly observed data only, and a bridge model using EU Economic Sentiment Indicator (ESI). Bridge models use a coincident indicator i_t aggregated to the quarterly level $i_t^{(Q)}$ to predict GDP y_t of the same quarter. To make forecasts, i_t is predicted with an ARIMA model. As in Itkonen and Juvonen (2017), we use two lags of aggregated indicators to predict GDP and automatic ARIMA model selection of Hyndman and Khandakar (2008) to forecast the indicator.

Table 5 shows the RMSFE of the GDP for the whole sample, and Table 6 shows them only for the pre-Covid sample ending

TABLE 5 | Out-of-sample forecasting results, full sample.

	Forecast horizon h						
Model	**0**	**1**	**2**	**3**	**4**	**5**	**6**
PLS-DFM ($p = 1, q = 1$)	1.455*	**1.288****	**1.421****	**1.471***	**1.573**	1.668	1.633
VAR-PLSR ($p = 1, q = 1$)	1.439*	1.677	1.582	1.620	1.655	1.696	1.625
ML-DFM ($p = 1, q = 1$)	1.588	1.592	1.609	1.617	1.620	1.628	1.628
PLS-DFM ($p = 3, q = 3$)	1.448*	1.365**	1.460*	1.542	1.644	1.710	1.623
VAR-PLSR ($p = 3, q = 3$)	1.551	1.785	1.657	1.641	1.723	1.747	1.615
ML-DFM ($p = 3, q = 3$)	1.574	1.576	1.587	1.602	1.617	**1.625**	1.623
PLS-DFM ($p = 1, \phi_r > 25\%$)	1.459*	1.329**	1.491*	1.541	1.601	1.646	1.627
VAR-PLSR ($p = 1, \phi_r > 25\%$)	1.461*	1.633	1.572	1.594	1.630	1.662	1.619
PLS-DFM ($p = 3, \phi_r > 25\%$)	**1.411***	1.359**	1.450*	1.553	1.616	1.689	1.611
VAR-PLSR ($p = 3, \phi_r > 25\%$)	1.528	1.760	1.651	1.654	1.690	1.720	**1.606**
Bridge	1.551	1.555	1.958	1.610	1.694	1.776	1.640
AR	1.638			1.623			1.627

Note: The entries are out-of-sample root mean squared forecast errors (RMSFEs). The first training sample starts in February 1995, and the test sample runs from January of 2007 to December of 2023. All methods use a rolling window of 137 observations and ML-DFM includes AR-dynamics for the idiosyncratic component. The forecast horizon h is the number of months until the end of the quarter that is being nowcasted or forecasted. The text in **bold** refers to the best method for each horizon and * (†) on PLS-based method (non-PLS based method) marks a PLS based (non-PLS based) method that has 5% smaller RMSFEs than any non-PLS based (PLS based) method for the same horizon. Double ** (††) refers to a 10% difference.

TABLE 6 | Out-of-sample forecasting results, pre-Covid sample.

	Forecast horizon h						
Model	**0**	**1**	**2**	**3**	**4**	**5**	**6**
PLS-DFM ($p = 1, q = 1$)	1.321	**1.213***	1.404	1.285	1.334	1.398	1.410
VAR-PLSR ($p = 1, q = 1$)	**1.244***	1.335	1.358	1.338	1.360	1.400	1.414
ML-DFM ($p = 1, q = 1$)	1.352	1.357	1.390	1.400	1.404	1.414	1.413
PLS-DFM ($p = 3, q = 3$)	1.319	1.228*	1.361	1.283	1.353	1.387	1.384
VAR-PLSR ($p = 3, q = 3$)	1.403	1.454	1.384	1.354	1.403	1.423	1.388
ML-DFM ($p = 3, q = 3$)	1.345	1.352	1.376	1.385	1.393	1.404	1.407
PLS-DFM ($p = 1, \phi_r > 25\%$)	1.392	1.263*	1.432	1.268*	1.325	1.368	1.391
VAR-PLSR ($p = 1, \phi_r > 25\%$)	1.294	1.304	1.366	1.304	1.326	1.373	1.398
PLS-DFM ($p = 3, \phi_r > 25\%$)	1.313	1.228*	1.365	**1.262***	**1.324**	1.358	**1.361**
VAR-PLSR ($p = 3, \phi_r > 25\%$)	1.388	1.440	1.396	1.343	1.374	1.405	1.370
Bridge	1.324	1.332	1.377	1.339	1.339	**1.294**	1.375
AR	1.394			1.404			1.409

Note: See Table 5.

in 2019Q4. We show results for these samples separately as the Covid-19 recession was a radical and extremely volatile unforeseeable event. Each column with different h reports the number of months until the end of the quarter that is being nowcasted or forecasted. For example, the nowcast made in March for the first quarter would get a value of $h = 0$ and the forecast for the next quarter would get a value of $h = 3$. Since the AR(1) model is used for quarterly GDP growth and changing data vintages are ignored, the results are only showed once per quarter. The best method for each forecast horizon is bolded.

For the full sample in Table 5, we see that the PLS-DFM models performs comparatively well up to horizon $h = 2$ regardless of the choice of q or p. At horizons $h = 3$ and $h = 4$, the smallest PLS-DFM seems to perform best. At horizons $h = 5$ and $h = 6$, all the methods perform very similarly. Meanwhile, the small

VAR-PLSR performs well at horizon $h = 0$, but is otherwise very similar to ML-DFM. Increasing the number of factors generally makes the PLS based methods perform worse (when comparing $q = 1$ and $q = 3$ cases). This suggests that there are also non-persistent factors in the Finnish economy that do not really benefit the factor model. In fact, the greater complexity harms the model. Choosing q by the decision rule $\phi_r > 25\%$ seldom results in the best model, but is still a good choice. Meanwhile, the performance of ML-DFM is not substantially affected by the addition of factors.

Table 6 gives the corresponding results for the pre-Covid sample period ending in 2019Q4. Here the differences between RMSFEs are a lot smaller. However, at most horizons PLS-based methods perform the best. The decision rule $\phi_r > 0.25$ is the best option at horizons $h = 2,3,4$, or 6. However, the simple bridge model performs the best at horizon $h = 5$. Additionally, we again observe that the PLS-based methods benefit from greater parsimony, whereas ML-DFM seems to always benefits from the addition of factors. Again, this would suggest that some factors are more persistent and useful for forecasting than others.

Tables 5 and 6 suggest that RMSFEs of PLS based methods increased less than those of ML-DFM during and after Covid-19. One possible reason for this robustness is that PLS focuses on more persistent structures in time series than ML (See Section 3). The main impact of Covid-19 on the (differenced) time series was strong but ultimately transitory. By not focusing on such events too strongly, PLS can stay more focused on the predictable developments of the response.

The Covid period that is included in Table 5 was a period of unprecedented volatility, and the differences between models in Table 6 were small. In the prior table the high volatility and in the latter table small differences may mean that the differences are not statistically significant. To check this, Tables 7 and 8 report Z-statistics for the Diebold-Mariano test (DM) for pairwise model comparisons at $h = 3$.[5] Each row and column represents a model, although model names are omitted from columns to save space. High (low) value for the statistic means that the row model is worse (better) than the column model.

The test statistics of Table 7 suggest that the small $p = 1$ and $q = 1$ PLS-DFM model is indeed better than many other models compared (the first column is positive). Likewise PLS-DFM is better than the comparable VAR-PLSR in the fixed q cases. In the pre-Covid world the comparisons are less conclusive.

In summary, the PLS-based methods perform better than ML-DFM, but this difference might not be statistically significant. This difference appears larger and clearer during turbulent times, such as the Covid recession. This would suggest that the PLS-based methods are more robust to such events. Furthermore, PLS performs better with fewer factors, whereas ML benefits from more factors. This suggests that even a small PLS-DFM can capture the persistent intertemporal dynamics present in the data. Also, even though the VAR-PLSR did not perform especially well in the simulations of Section 4, it appears to perform rather well at short forecast horizons with real-world data.

5.2 | Comparing the Estimated Factors

In Figure 2, we have estimated three DFMs for the Finnish macroeconomic data of Section 5.1, each with $q = p = 1$. The left panel shows cumulative factor estimates, that is, $\sum_{\tau=0}^{t} f_\tau$. These start from zero and end in zero because the data is centered. All three curves have a very similar shape, and they seem to generally capture the business cycles of the Finnish economy from February

TABLE 7 | Full sample horizon $h = 3$ DM test statistics.

		a	b	c	d	e	f	g	h	i	j	k
a	PLS-DFM ($p = 1, q = 1$)											
b	VAR-PLSR ($p = 1, q = 1$)	1.77*										
c	ML-DFM ($p = 1, q = 1$)	1.48^{+}	−0.03									
d	PLS-DFM ($p = 3, q = 3$)	0.84	−1.48^{+}	−0.64								
e	VAR-PLSR ($p = 3, q = 3$)	1.27	0.30	0.16	−1.72*							
f	ML-DFM ($p = 3, q = 3$)	1.42^{+}	−0.16	−2.17**	0.54	−0.26						
g	PLS-DFM ($p = 1, \phi_r > 25\%$)	0.84	−1.74*	−0.77	−0.03	−1.36^{+}	−0.65					
h	VAR-PLSR ($p = 1, \phi_r > 25\%$)	1.09	−0.58	−0.18	0.99	−0.97	−0.07	1.31^{+}				
i	PLS-DFM ($p = 3, \phi_r > 25\%$)	0.76	−1.10	−0.48	0.43	−2.13**	−0.38	0.28	−0.94			
j	VAR-PLSR ($p = 3, \phi_r > 25\%$)	1.14	0.38	0.23	1.37^{+}	0.41	0.32	1.26	1.02	1.69*		
k	Bridge	1.01	0.10	0.05	0.98	0.39	0.05	0.72	0.16	0.85	0.49	
l	AR	1.49^{+}	0.02	0.97	0.68	0.12	1.83*	0.81	0.23	0.52	0.19	0.09

Note: The entries are Z-statistics for the Diebold-Mariano test at $h = 3$. Rows and columns have different models in the same order. The model names are omitted from columns to save space, but letters are added to increase readability. Positive values mean that the model on the row is worse than the one on the column. See Table 5 for additional notes.
Symbol ** refers to statistical significance at 5%, * at 10%, and $^{+}$ at 20% (10% one-sided).

TABLE 8 | Pre-Covid horizon $h = 3$ DM test statistics.

		a	b	c	d	e	f	g	h	i	j	k
a	PLS-DFM $(p = 1, q = 1)$											
b	VAR-PLSR $(p = 1, q = 1)$	−1.02										
c	ML-DFM $(p = 1, q = 1)$	0.91	−0.38									
d	PLS-DFM $(p = 3, q = 3)$	0.02	0.80	0.67								
e	VAR-PLSR $(p = 3, q = 3)$	−0.58	−0.18	0.22	−1.12							
f	ML-DFM $(p = 3, q = 3)$	−0.85	−0.31	1.63^{+}	−0.62	−0.15						
g	PLS-DFM $(p = 1, \phi_r > 25\%)$	0.66	1.31^{+}	0.95	0.26	0.82	0.89					
h	VAR-PLSR $(p = 1, \phi_r > 25\%)$	−0.30	1.01	0.57	−0.33	0.67	0.51	−0.70				
i	PLS-DFM $(p = 3, \phi_r > 25\%)$	0.26	0.99	0.73	1.19	1.67*	0.68	0.08	0.64			
j	VAR-PLSR $(p = 3, \phi_r > 25\%)$	−0.48	−0.05	0.27	−0.92	0.92	0.21	−0.71	−0.53	-1.46^{+}		
k	Bridge	−0.33	0.00	0.28	−0.53	0.13	0.22	−0.48	−0.23	−0.75	0.04	
l	AR	−0.90	−0.40	−0.52	−0.68	−0.23	−1.23	−0.94	−0.57	−0.73	−0.28	−0.29

Note: See Table 7.

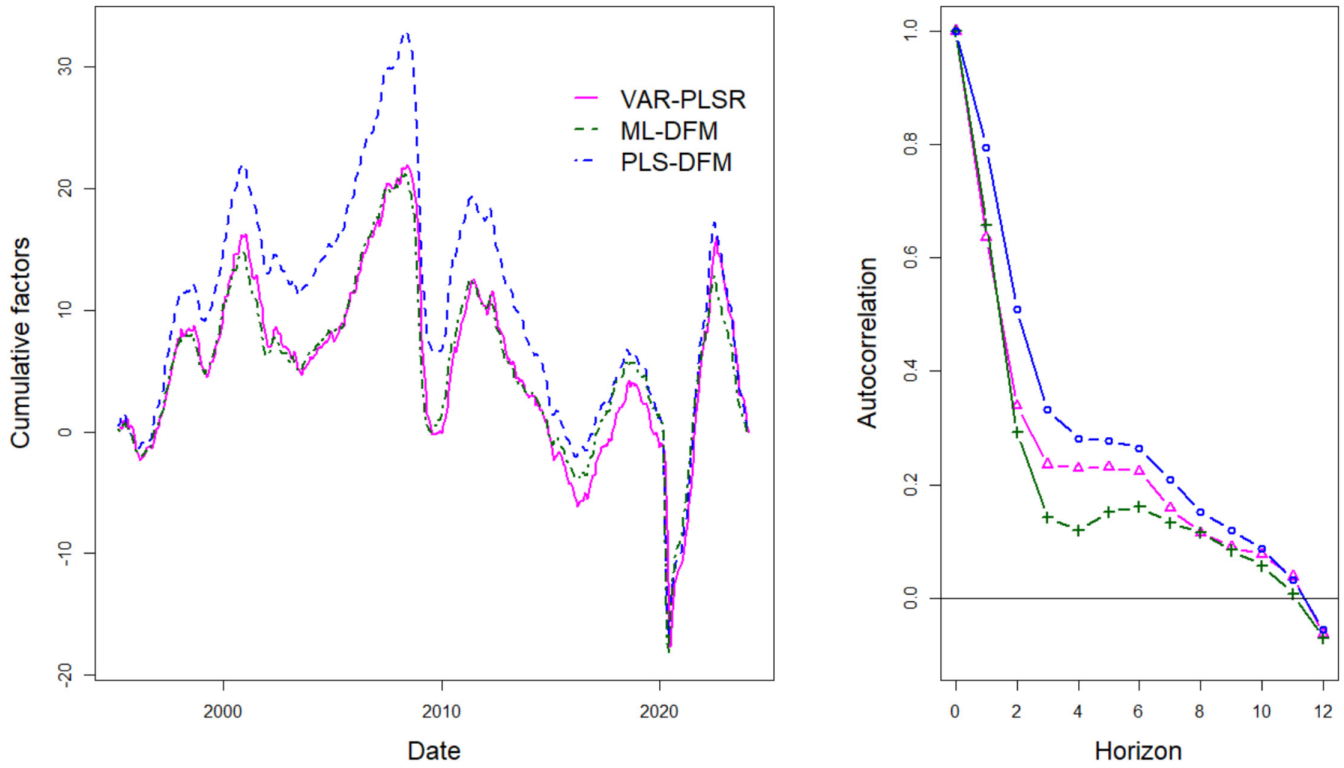

FIGURE 2 | Cumulative factor estimates (left) and factor autocorrelations (right).

1995 to December 2023. However, the PLS-DFM estimate is almost everywhere higher. This means that the cumulative factor of the PLS-DFM grows more in the early sample and less in the late sample than the other two methods. In other words, the PLS-DFM emphasizes the difference between the fast economic growth in Finland between the depression of the early 1990s and the global financial crisis and the slower growth since. This difference can be seen as a very subtle and highly persistent shift.

The right panel in Figure 2 shows the sample autocorrelation coefficients of the factors of different methods. The PLS-based methods result in factors that are more persistent than those obtained with the ML-based approach. However, these differences are quite small. At the 95% confidence level, the difference between the autocorrelations of PLS-DFM and ML-DFM is statistically significant only at lag 2 (testing not shown in Figure 2).

6 | Conclusions

This paper proposes a way of estimating a dynamic factor model using partial least squares, the PLS-DFM. Unlike PCA or ML, PLS uses information about the forecast target in the factor estimation procedure. This allows for more parsimonious model specifications that can still capture persistent dynamics that are useful in forecasting. Additionally, the approach is relatively simple, computationally well applicable and able to deal with important specific features of macroeconomic time series such as missing observations, ragged edge and mixed frequency data. In addition to the PLS-DFM, a novel vector autoregressive PLSR (VAR-PLSR) was also investigated. The matter of estimating the number of persistent factors was also investigated in Sections 2.2 and 4.4.

Simulation results generally suggest comparable performance by the method, when compared to the maximum likelihood based approach (ML-DFM) and superior when compared to the commonly used two-step estimator. However, when the DGP has a mix of persistent and non-persistent factors, and we are only interested in the persistent ones, PLS-DFM generally gave the best performance. Likewise, PLS seems to perform better with mixed frequency data in small samples. Our empirical forecasting results obtained for the Finnish GDP growth show that the PLS-DFM is often superior to the ML-DFM although the difference is not always highly statistically significant. Meanwhile, the VAR-PLSR also proved useful in the empirical forecasting application.

Acknowledgments

The author would like to thank Petteri Juvonen, Juho Koistinen, Markku Lanne, Henri Nyberg, Joni Virta, and seminar participants at the Bank of Finland (2024), Nordic Econometric Meeting (Bergen 2024), CFE 2024 (London), and Helsinki Graduate School of Economics (2024). The financial support from the OP Group Research Foundation (grant 20230116), the Foundation for Economic Education (Liikesivistysrahasto, grant 220246), the Turku University Foundation (Turun Yliopistosäätiö, grant 081875), and The Finnish Doctoral Program Network in Artificial Intelligence, AI-DOC (decision number VN/3137/2024-OKM-6) is gratefully acknowledged. Special thanks are due to the Bank of Finland for the data used in Section 5 and research cooperation (research visit in the autumn 2024). Open access publishing facilitated by Turun yliopisto, as part of the Wiley - FinELib agreement.

Ethics Statement

The author has nothing to report.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in Statistics Finland at https://stat.fi/tup/index_en.html.

Endnotes

[1] Kalman filter is a key feature of many DFMs as it can improve our estimations of the factors themselves, make it easier to include autocorrelated idiosyncratic components and it helps dealing with missing observations or mixed frequency data. The problem of missing data, and especially the so called "ragged edge," that is, some variables are observed only with a significant lag (see, for example, Giannone et al. 2008), is at the heart of nowcasting.

[2] To clarify: X and Y are $T \times k$ and $T \times n$-dimensional matrices respectively, v_r and u_r are T- dimensional vectors, λ_{xr} and λ_{yr} are k and n- dimensional vectors respectively, p_r is also a k-dimensional vector and d_r is a scalar.

[3] Essentially $\text{Cor}(\xi_{it}, w_t) \to 0$ when $n \to \infty$ for all $i = 1, \ldots, n$.

[4] Due to there being just one persistent factor, the trace operator does not matter. Also, this statistic is in practice just the regular R^2 of a regression between $\widehat{F}_{T|T}$ and F_T.

[5] Testing is carried out by linear regression on the difference of squared errors with Newey-West HAC-robust standard errors as suggested by Diebold (2015). The construction of standard errors uses Bartlett kernel with lag of 3.

References

Bańbura, M., and M. Modugno. 2014. "Maximum Likelihood Estimation of Factor Models on Datasets With Arbitrary Pattern of Missing Data." *Journal of Applied Econometrics* 29, no. 1: 133–160.

Boivin, J., and S. Ng. 2006. "Are More Data Always Better for Factor Analysis?" *Journal of Econometrics* 132, no. 1: 169–194.

Chamberlain, G., and M. Rothschild. 1983. "Arbitrage, Factor Structure, and Mean-variance Analysis on Large Asset Markets." *Econometrica* 51, no. 5: 1281.

Connor, G., and R. Korajczyk. 1986. "Performance Measurement With the Arbitrage Pricing Theory." *Journal of Financial Economics* 15: 373–394.

Diebold, F. X. 2015. "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold Mariano Tests." *Journal of Business & Economic Statistics* 33, no. 1: 1–1.

Doz, C., D. Giannone, and L. Reichlin. 2011. "A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering." *Journal of Econometrics* 164, no. 1: 188–205.

Doz, C., D. Giannone, and L. Reichlin. 2012. "A Quasi-Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models." *Review of Economics and Statistics* 94, no. 4: 1014–1024.

Eickmeier, S., and T. Ng. 2011. "Forecasting National Activity Using Lots of International Predictors: An Application to New Zealand." *International Journal of Forecasting* 27, no. 2: 496–511.

Engle, R. F., and M. W. Watson. 1983. "Alternative Algorithms for the Estimation of Dynamic Factor, Mimic and Varying Coefficient Regression Models." *Journal of Econometrics* 23, no. 3: 385–400.

Engle, R., and M. Watson. 1981. "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates." *Journal of the American Statistical Association* 76, no. 376: 774–781.

Forni, M., and L. Reichlin. 1996. "Dynamic Common Factors in Large Cross-Sections." *Empirical Economics* 21: 27–42.

Forni, M., and L. Reichlin. 1998. "Lets Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycle." *Review of Economic Studies* 65: 453–474.

Fuentes, J., P. Poncela, and J. Rodríguez. 2015. "Sparse Partial Least Squares in Time Series for Macroeconomic Forecasting." *Journal of Applied Econometrics* 30, no. 4: 576–595.

Geweke, J. 1977. "The Dynamic Factor Analysis of Economic Time Series." In Latent Variables in Socio-Economic Models.

Giannone, D., L. Reichlin, and D. Small. 2008. "Nowcasting: The Real-Time Informational Content of Macroeconomic Data." *Journal of Monetary Economics* 55, no. 4: 665–676.

Groen, J. J. J., and G. Kapetanios. 2016. "Revisiting Useful Approaches to Data-rich Macroeconomic Forecasting." *Computational Statistics & Data Analysis* 100: 221–239.

Helland, I. S. 1990. "Partial Least Squares Regression and Statistical Models." *Scandinavian Journal of Statistics*: 97–114.

Hepenstrick, C., and M. Marcellino. 2019. "Forecasting Gross Domestic Product Growth With Large Unbalanced Data Sets: The Mixed Frequency Three-pass Regression Filter." *Journal of the Royal Statistical Society Series A: Statistics in Society* 182, no. 1: 69–99.

Hindrayanto, I., S. J. Koopman, and J. de Winter. 2016. "Forecasting and Nowcasting Economic Growth in the Euro Area Using Factor Models." *International Journal of Forecasting* 32, no. 4: 1284–1305.

Höskuldsson, A. 1988. "PLS Regression Methods." *Journal of Chemometrics* 2, no. 3: 211–228.

Hyndman, R. J., and Y. Khandakar. 2008. "Automatic Time Series Forecasting: The Forecast Package for R." *Journal of Statistical Software* 27: 1–22.

Itkonen, J., and P. Juvonen. 2017. "Nowcasting the Finnish Economy With a Large Bayesian Vector Autoregressive Model." BoF Economics Review 6/2017.

Jungbacker, B., and S. J. Koopman. 2015. "Likelihood-Based Dynamic Factor Analysis for Measurement and Forecasting." *Econometrics Journal* 18, no. 2: C1–C21.

Juvonen, P., and A. Lindblad. 2025. "Nowcasting in Real Time: Large Bayesian Vector Autoregression in a Test (No. 6/2025)." Bank of Finland Research Discussion Papers.

Kelly, B., and S. Pruitt. 2013. "Market Expectations in the Cross-Section of Present Values." *Journal of Finance* 68, no. 5: 1721–1756.

Kelly, B., and S. Pruitt. 2015. "The Three-Pass Regression Filter: A New Approach to Forecasting Using Many Predictors." *Journal of Econometrics* 186, no. 2: 294–316.

Krantz, S., and R. Bagdziunas. 2023. "dfms: Dynamic Factor Models." R package version 0.2.1. https://CRAN.R-project.org/package=dfms.

Lohmöller, J. B. 2013. *Latent Variable Path Modeling With Partial Least Squares*. Springer Science & Business Media.

Mariano, R. S., and Y. Murasawa. 2003. "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series." *Journal of Applied Econometrics* 18, no. 4: 427–443.

Ng, S. K., T. Krishnan, and G. J McLachlan. 2011. "The EM Algorithm." *In Handbook of Computational Statistics: Concepts and Methods*, 139–172.

Sargent, T., and C. A. Sims. 1977. "Business Cycle Modeling Without Pretending to Have Too Much a Priori Economic Theory." In *In New Methods in Business Cycle Research: Proceedings From a Conference*, Federal Reserve Bank of Minneapolis, 45–109.

Stock, J. H., and M. W. Watson. 1989. "New Indexes of Coincident and Leading Economic Indicators." NBER Macroeconomics Annual 4.

Stock, J. H., and M. W. Watson. 2002. "Forecasting Using Principal Components From a Large Number of Predictors." *Journal of the American Statistical Association* 97, no. 460: 1167–1179.

Stock, J. H., and M. W. Watson. 2016. "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics." *In Handbook of Macroeconomics*, 2. Elsevier, 415–525.

Stundziene, A., V. Pilinkiene, J. Bruneckiene, A. Grybauskas, M. Lukauskas, and I. Pekarskiene. 2024. "Future Directions in Nowcasting Economic Activity: A Systematic Literature Review." *Journal of Economic Surveys* 38, no. 4: 1199–1233.

Wold, H. 1975. "Path Models With Latent Variables: The NIPALS Approach." *In Quantitative Sociology*. Academic Press, 307–357.

Appendix A

Comparison of Computation Times

In this appendix, we show the average computation times in the simulations of Section 4. The settings here are the same as in Tables 1 and 3. However, Panels D and G are omitted, since the ML-DFM estimates more parameters, and the comparison would not be fair. Also, we use a conventional ML procedure described in detail in Engle and Watson (1983) and Bańbura and Modugno (2014) (cf. Jungbacker and Koopman 2015). We use the Kalman smoother implementation of the R-package "dfms" by Krantz and Bagdziunas (2023) and the eigenvalue decomposition implementation of base R. Otherwise, the code is our own. Computation times only include the training of the model and not the time it takes to construct forecasts. All computations are carried out on the same computer, and ML is initiated by the two-step estimate. EM-algorithm is deemed converged when the log-likelihood increases by less than 0.01 and PLS is deemed converged when the mean absolute variation of u_r of Algorithms 1 and 2 between iterations is less than 0.0001.

The results are shown in Table A1. The results show that the PLS based methods are roughly q times slower than 2s-DFM. This is likely because the PLS algorithm estimates the factor loadings one at a time. Nevertheless, both 2s-DFM and the PLS based methods are fairly quick. Meanwhile, ML-DFM is not much slower than PLS-DFM when n is small, but as it grows, computation times of ML-DFM can be an order of magnitude longer than those of PLS-DFM. This was the main challenge that faced ML in the past. Additionally, computation times for ML-DFM are higher in Panels E and F than in Panels A–C even though the number of estimated factors is lower. This is unique to ML-DFM and suggests that the EM-algorithm is slowed by non-persistent factors.

TABLE A1 | Computation times for different DGPs.

Panel A: $\rho = 0.9, \alpha = 0, \tau = 0, u = 0.1, q = \widehat{q} = 1$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.001	0.007	0.036	0.016	0.070	0.305	0.004	0.008	0.014	0.005	0.012	0.036
100	0.002	0.012	0.071	0.021	0.098	0.502	0.007	0.011	0.018	0.010	0.018	0.062

Panel B: $\rho = 0.9, \alpha = 0, \tau = 0, u = 0.1, q = \widehat{q} = 3$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.002	0.008	0.040	0.068	0.154	0.513	0.032	0.061	0.098	0.033	0.066	0.123
100	0.002	0.014	0.077	0.072	0.199	0.747	0.050	0.101	0.159	0.054	0.110	0.209

Panel C: $\rho = 0.7, \alpha = 0, \tau = 0, u = 0.1, q = \widehat{q} = 3$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.002	0.007	0.040	0.073	0.145	0.474	0.037	0.077	0.127	0.038	0.082	0.152
100	0.002	0.014	0.076	0.086	0.198	0.722	0.058	0.118	0.188	0.061	0.127	0.238

Panel E: $\rho = 0.9, \alpha = 0, \tau = 0, u = 0.1, q = 3, \widehat{q} = 1$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.002	0.007	0.037	0.057	0.246	1.065	0.009	0.02	0.032	0.010	0.024	0.055
100	0.002	0.012	0.070	0.080	0.484	2.563	0.014	0.02	0.032	0.017	0.027	0.075

Panel F: $\rho = 0.7, \alpha = 0, \tau = 0, u = 0.1, q = 3, \widehat{q} = 1$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.002	0.007	0.038	0.042	0.171	0.853	0.011	0.023	0.038	0.012	0.027	0.061
100	0.002	0.013	0.072	0.062	0.371	2.347	0.016	0.024	0.038	0.019	0.032	0.083

Note: The entries are the average computation times in seconds. See Tables 1 and 3 for details.

Finally, the difference between PLS-DFM and VAR-PLSR is mostly due to Kalman filter.

Appendix B

Additional Simulations

In this appendix, we describe additional simulation results concerning higher forecast horizons and mixed frequency data in the mixed persistent and non-persistent factor case.

Table A2 portrays results for DGPs of Section 4, where the forecast horizon is set to $h = 3$ instead of $h = 1$ as is done in Section 4. The results suggest that as forecast horizon grows, all the forecasting methods become less accurate. However, the relative performance between the four methods is the same as in Table 1.

Tables A3 and A4 are mixed frequency counterparts of Table 3 and 4 for mixed frequency data respectively. Like in the Section 4.2, Table A3 only reports RMSFEs for the quarterly variables.

TABLE A2 | Forecast simulation results: Higher forecast horizon $h = 3$.

Panel A: $\rho = 0.9, \alpha = 0, \tau = 0, u = 0.1, q = \widehat{q} = 1$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.847	0.849	0.852	0.806	0.810	0.813	0.85	0.849	0.851	0.816	0.812	0.814
100	0.837	0.833	0.837	0.792	0.789	0.792	0.84	0.833	0.837	0.800	0.791	0.793

Panel B: $\rho = 0.9, \alpha = 0, \tau = 0, u = 0.1, q = \widehat{q} = 3$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.884	0.875	0.874	0.845	0.836	0.834	0.870	0.855	0.855	0.860	0.839	0.836
100	0.864	0.848	0.849	0.809	0.798	0.802	0.851	0.837	0.837	0.828	0.804	0.804

Panel C: $\rho = 0.7, \alpha = 0, \tau = 0, u = 0.1, q = \widehat{q} = 3$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.996	0.994	0.994	0.991	0.983	0.981	0.995	0.990	0.991	0.990	0.982	0.982
100	0.995	0.997	0.998	0.977	0.977	0.979	0.991	0.994	0.994	0.981	0.978	0.979

Panel D: $\rho = 0.9, \alpha = 0.5, \tau = 0.5, u = 0.1, q = \widehat{q} = 3$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.893	0.880	0.879	0.841	0.832	0.833	0.865	0.856	0.858	0.848	0.839	0.838
100	0.871	0.849	0.851	0.816	0.798	0.802	0.855	0.838	0.840	0.824	0.805	0.807

Panel E: $\rho = 0.9, \alpha = 0, \tau = 0, u = 0.1, q = 3, \widehat{q} = 1$

	2s-DFM			ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n			n		
T	10	50	100	10	50	100	10	50	100	10	50	100
50	0.988	0.990	0.988	0.979	0.987	0.985	0.957	0.951	0.952	0.953	0.943	0.943
100	0.990	0.991	0.991	0.978	0.986	0.988	0.951	0.944	0.944	0.944	0.932	0.931

Note: Selected DGPs with the forecast horizon $h = 3$. See Table 1 for details.

TABLE A3 | Forecast simulation results: Mixed frequency and mix of persistent and non-persistent factors.

Panel E1m: $\rho = 0.9, \alpha = 0, \tau = 0, u = 0.1, q = 3, \widehat{q} = 1$

	ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n		
T	10	50	100	10	50	100	10	50	100
51	0.956	0.955	0.956	0.891	0.896	0.901	0.794	0.787	0.810
99	0.973	0.973	0.973	0.875	0.905	0.901	0.762	0.767	0.785

Panel F1m: $\rho = 0.7, \alpha = 0, \tau = 0, u = 0.1, q = 3, \widehat{q} = 1$

	ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n		
T	10	50	100	10	50	100	10	50	100
50	0.974	0.971	0.971	0.975	0.975	0.970	0.894	0.879	0.892
100	0.984	0.983	0.983	0.946	0.970	0.971	0.874	0.871	0.884

Panel G1m: $\rho = 0.9, \alpha = 0.5, \tau = 0.5, u = 0.1, q = 3, \widehat{q} = 1$

	ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n		
T	10	50	100	10	50	100	10	50	100
51	0.858	0.852	0.855	0.900	0.904	0.91	0.796	0.795	0.819
99	0.924	0.913	0.915	0.885	0.912	0.91	0.765	0.776	0.795

Note: In the table (Panels E1–G1), only the first factor is estimated and only one factor is persistent (non-zero autoregressive coefficients) in the DGP. See Table 1 for details.

TABLE A4 | Factor estimation simulation: Mixed frequency and mix of persistent and non-persistent factors.

Panel E2m: $\rho = 0.9, \alpha = 0, \tau = 0, u = 0.1, q = 3, \widehat{q} = 1$

	ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n		
T	10	50	100	10	50	100	10	50	100
50	0.232	0.202	0.183	0.482	0.595	0.613	0.536	0.610	0.615
100	0.293	0.265	0.252	0.639	0.776	0.796	0.710	0.797	0.807

Panel F2m: $\rho = 0.7, \alpha = 0, \tau = 0, u = 0.1, q = 3, \widehat{q} = 1$

	ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n		
T	10	50	100	10	50	100	10	50	100
50	0.185	0.190	0.187	0.541	0.708	0.738	0.582	0.712	0.729
100	0.193	0.195	0.190	0.640	0.835	0.866	0.700	0.850	0.869

Panel G2m: $\rho = 0.9, \alpha = 0.5, \tau = 0.5, u = 0.1, q = 3, \widehat{q} = 1$

	ML-DFM			VAR-PLSR			PLS-DFM		
	n			n			n		
T	10	50	100	10	50	100	10	50	100
50	0.229	0.209	0.191	0.493	0.597	0.616	0.546	0.613	0.618
100	0.279	0.264	0.257	0.643	0.778	0.797	0.711	0.798	0.808

Note: The entries are the trace R^2_{trace} statistics for the first (persistent) factor. See Table 4. Selected DGPs have 20% quarterly frequency variables and 80% monthly. Forecast horizon $h = 3$. See Table 1 for details.

Appendix C

Data Description

We thank Petteri Juvonen of the Bank of Finland for compiling the data set, which is originally from Statistics Finland. Our data set is the same as in Juvonen and Lindblad (2025), except that we have omitted the number of overnight stays at accommodations, number of bankruptcy cases and the consumer confidence in the overall economy (as opposed to "Consumer survey: Own economy"). These changes were made due to the Covid-19 pandemic. We use first differences of all series. The data is described in Table A5.

TABLE A5 | Data description.

	Frequency	Log	Lag
Gross domestic product	Q	X	2
Private consumption expenditure	Q	X	2
Government consumption expenditure	Q	X	2
Gross fixed capital formation, residential buildings	Q	X	2
Gross fixed capital formation, excluding residential buildings	Q	X	2
Exports of goods and services	Q	X	2
Imports of goods and services	Q	X	2
Index of wage and salary earnings	Q	X	1
Price index of dwellings	Q	X	2
Volume index of industrial output	M	X	1
Capacity utilization rate, Manufacturing	M		1
Granted building permits	M	X	2
Turnover of retail trade, volume index	M	X	1
Turnover of wholesale trade, volume index	M	X	1
Turnover of motor vehicle trade, volume index	M	X	1
Manufacturing working on orders, Index of turnover in industry	M	X	2
Exports of goods	M	X	1
Imports of goods	M	X	1
Employed, ages 15–74	M	X	1
Unemployment rate, ages 15–74	M		1
Jobs vacant	M	X	1
Unemployed job seekers	M	X	1
OMXHelsinki All-Share Index	M	X	0
Consumer survey: Own economy	M		0
Business confidence, manufacturing	M		0
Business confidence, construction	M		0
Business confidence, manufacturing: production expectations	M		0
Consumer price index	M	X	0
Building cost index	M	X	2
Turnover of construction, volume index	M	X	3
Producer price index, manufacturing	M	X	1
Export price index	M	X	1
Import price index	M	X	1
New orders in manufacturing	M	X	1
ISM's Manufacturing PMI Index, USA	M		0

(Continues)

TABLE A5 | (Continued)

	Frequency	Log	Lag
ifo Business Climate Index, Germany	M		0
Wages and salaries sum	M	X	1
Volume index of new building	M	X	2
Economic sentiment indicator, Eurozone	M		0
Index of turnover in industry	M	X	3
Building starts	M	X	3
Building completions	M	X	3
World trade	M	X	3
Turnover of service industries	M	X	2

Note: "Frequency" refers to the observation frequency: "Q" means that the observed frequency is quarterly, and "M" means that the frequency is monthly. If "Log" is X, it means that we take the logarithmic transformation of the series before differencing. Finally, "Lag" is the publication lag in months (end-of-month).