Journal of Econometrics 193 (2016) 405–417

Contents lists available at ScienceDirect

Journal of Econometrics

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

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

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

[a] Vrije Universiteit Amsterdam, Netherlands

[b] Tinbergen Institute Amsterdam, Netherlands

[c] CREATES, Aarhus University, Denmark

ARTICLE INFO

Article history:
Available online 6 May 2016

JEL classification:
C13
C32
C53
E17

Keywords:
Asymptotic theory
Kalman filter
Nowcasting
State space

ABSTRACT

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

1. Introduction

The forecasting of macroeconomic and financial time series variables is of key importance for economic policy makers. Reliable forecasts are especially in high demand when the economic environment is uncertain as we have witnessed in the years during and after the financial crisis. Many different model-based approaches exist for this purpose, ranging from basic time series models to sophisticated structural dynamic macroeconomic models. The underlying idea of the dynamic factor model is to associate a relatively small set of factors to a high-dimensional panel of economic variables that includes the variables of interest and related variables. The dynamic factor model has become a popular tool for the forecasting of the variable of interest, amongst practitioners and econometricians. This is mainly due to their good forecast performance as shown in many studies.

The dynamic factor model can be viewed as a high-dimensional linear state space model. The estimation of the parameters in a dynamic factor model is a challenging task given the large number of parameters, mostly due to factor loading coefficients. A likelihood-based approach in which the Gaussian likelihood function is evaluated via the Kalman filter and is numerically maximized with respect to the parameter vector has been originally proposed by Engle and Watson (1981) for a model with one dynamic factor. Watson and Engle (1983) base their estimation procedure on an expectation–maximization (EM) algorithm; see also Quah and Sargent (1993). More recently, feasible two-step approximate likelihood-based procedures are developed by Doz et al. (2011) and Bańbura and Modugno (2014). In Bräuning and Koopman (2014) and Jungbacker and Koopman (2015), specific data transformations are considered to facilitate the parameter estimation for high-dimensional dynamic factor models. In this study, we restrict ourselves to likelihood-based estimation procedures.

* Correspondence to: Department of Econometrics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081HV Amsterdam, The Netherlands.

E-mail addresses: f.blasques@vu.nl (F. Blasques), s.j.koopman@vu.nl (S.J. Koopman), maxmallee@hotmail.com (M. Mallee), z2.zhang@vu.nl (Z. Zhang).

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

Weighted likelihood-based estimation

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

The variable-specific weights introduced by our weighted ML estimator differ from other weighted ML estimators proposed in the literature. In most other cases, observation-specific weights in the likelihood function are considered. The local ML estimators studied in Tibshirani and Hastie (1987), Staniswalis (1989) and Eguchi and Copas (1998) assign a weight to each observation that depends on the distance to a given fixed point. The robust ML estimator of Markatou et al. (1997, 1998) down-weights observations that are inconsistent with the postulated model. Similarly, Hu and Zidek (unpublished) devise a general principle of relevance that assigns different weights to different observations in an ML setting. In small samples, this type of estimator can provide important gains in the trade-off between bias and precision of the ML estimator. The large sample properties of these estimators are established in Wang et al. (2004) for given weights, and Wang and Zidek (2005) provide a method for estimating the weights based on cross-validation. In contrast we propose a weighted ML estimator that gives higher weight to a subset of a random vector, that is to an entire random scalar sequence within the multivariate stochastic process.

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

Mixed Frequency

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

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

Empirical study

An important application of dynamic factor models is their use in the forecasting of quarterly GDP growth. A high-dimensional panel of macroeconomic variables is used to construct factors for the purpose of facilitating the forecasting of GDP growth. Empirical evidence is given by, amongst others, Stock and Watson (2002b) and Giannone et al. (2008) for the U.S., Marcellino et al. (2003) and Rünstler et al. (2009) for the euro area, and Schumacher and Breitung (2008) for Germany. In many of these and related studies, the problem of mixed frequency data arises since the variable of interest GDP growth is observed at a quarterly frequency while the other macroeconomic variables are observed at a monthly frequency. The treatment of mixed frequency data in a dynamic factor model is therefore a highly relevant issue in forecasting, nowcasting and backcasting GDP growth; see also the discussions in Bańbura et al. (2013).

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

Outline

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

2. Weighted maximum likelihood: method and properties

We represent our high-dimensional panel of time series as the column vector z_t for which we have observations from $t = 1, \dots, T$ where T is the overall time series length. We decompose z_t into variables of interest in y_t and related variables in x_t, we have $z_t = (y_t', x_t')'$ where a_t' is the transpose of column vector a_t. The dimension N_y of y_t is small and typically equal to one while the