Long Term Forecasting of El Niño Events via Dynamic Factor Simulations

Mengheng Li[(a)] *Siem Jan Koopman*[(a,b)*] *Rutger Lit*[(a)]

Desislava Petrova[(c)]

[(a)]Vrije Universiteit Amsterdam and Tinbergen Institute, The Netherlands

[(b)]CREATES, Aarhus University, Denmark

[(c)]Catalan Institute for Climate Science (IC3), Barcelona, Spain

May 2, 2017

*Corresponding author: S.J. Koopman, Department of Econometrics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. s.j.koopman@vu.nl

Long Term Forecasting of El Niño Events via Dynamic Factor Simulations

by M. Li, S.J. Koopman, R. Lit & D. Petrova

Abstract

We propose a new forecasting procedure for the Niño3.4 time series that is linked with the well-known El Niño phenomenon. This important climate time series is subject to an intricate serial correlation structure and is related to many other relevant and related variables. Although the forecasting procedure is valid for all lead times, it is particularly developed for medium to long term forecasting of El Niño. The procedure consists of three phases and relies on the subsequent use of an univariate time series method for producing prediction errors, the formulation of a dynamic factor model for these prediction errors and the explanatory variables, the simulation of many signal paths from the dynamic factor model conditional on all explanatory variables, and the reconstruction of artificial univariate time series of the variable of interest for which forecasts can be generated. The sample average of these ensemble forecasts is our final forecast. We justify each step of the procedure in general terms and relate the procedure to well-known concepts in econometrics. Our proposed forecasting procedure is specially developed for the important and challenging problem of forecasting the warm El Niño events. We provide evidence that our procedure is superior in the forecasting of El Niño when compared to other econometric forecasting methods.

Some key words: Climate research, Dynamic models, Kalman filter, Simulation smoothing.

JEL classification: C32, C42.

1 Introduction

El Niño is the important and well-known phenomenon of having higher than average sea surface temperatures in the central and eastern equatorial Pacific. It has an enormous impact on the climate in many parts of the world. Therefore, it has been given much coverage in the popular media, and it is the subject of extensive research in the scientific world. The occurrence of El Niño typically causes changes in weather patterns related to temperature, pressure and rainfall. An El Niño event may not only have a negative impact on local economies, but can also have negative consequences for public health, as in some regions these changes increase substantially the risk of water-borne and vector-borne diseases of which dengue is an example. Given its huge impact especially on developing countries in the proximity of the Pacific Ocean, it is self-evident that a timely forecast of the next El Niño event is of crucial importance. Much scientific research has been devoted to the development of forecasting methods for El Niño. The oscillation is characterized by an irregular period of between 2 and 7 years. Currently, forecasts are issued on a regular basis for up to three seasons in advance, but the long term of more than one year ahead forecasts remain a real challenge. Only in a few theoretical studies such long term forecasts are documented. At the same time the physics underlying El Niño implies that it is a self-sustaining climatic fluctuation that is quasi-periodic, with several dominant peaks in its spectrum, the main one being at about every 4-5 years, a secondary at about 2 years, and a third one at about 1.5 years. This suggests that it may be predictable at lead times on the order of several years. In this paper we propose a new forecasting procedure for El Niño (for the Niño3.4 time series, in particular) aimed at medium to long lead times. The forecasting methods are based on state-of-the-art developments in time series econometrics including state space models and simulation methods.

In this paper we address the classical forecasting problem in time series analysis where we need to forecast a single time series of interest that is subject to a possibly intricate serial correlation structure and for which a possibly large set of explanatory variables is available to us. The standard approach is to consider a linear model that simultaneously treats the dynamic structure in the time series and the explanatory variables which may be endogenous or exogenous. Both in econometrics and statistics, much attention is given

to the selection of the appropriate set of explanatory variables. The adopted methodology depends on whether the purpose is in-sample fit or out-of-sample prediction. In the latter case, it is typically argued that more parsimonious models are more successful in forecasting. There is a vast literature on forecasting in this setting and in the main paper we provide an overall discussion with appropriate references to the key forecasting methodologies in this setting and we will indicate their relevance to our novel approach.

We denote the variable of interest by the scalar y_t and the explanatory variables by the vector X_t for which we have observations at time points $t = 1, \ldots, T$ but we do not have observations available at "future" time points $T + 1, T + 2, \ldots$. We assume that the time series y_t and those in X_t are stationary but persistent processes. We abstract ourselves from non-stationary and cointegration issues in the current study. We further notice that X_t may also contain lagged explanatory variables. The aim is to produce an estimate for y_{T+h} where h is the forecast horizon, based on the available observations for y_t and X_t up to time T. For small values of h, we refer to the forecast made at time T, denoted by $\hat{y}_{T+h|T}$, as a *short term forecast* while for larger values of h we refer to the forecast as a *long term forecast*. It depends on the time series and the purpose of the study of what forecast horizon is associated with short, medium or long term forecasts. Our focus is mostly directed towards long term forecasting but our proposed procedure is also valid for short term forecasting. To produce a forecast for y_{T+h}, the information in $X_{T+1}, \ldots, X_{T+h}$ may be highly relevant for it but we do not have their observations available at time T. A possible solution to this problem is to produce forecasts for the explanatory vector $X_{T+1}, \ldots, X_{T+h}$ in an ad-hoc way from which the actual forecast $\hat{y}_{T+h|T}$ can be computed. Given that these forecasts for the explanatory variables may be inaccurate, especially for larger values of h, the forecast $\hat{y}_{T+h|T}$ is often less accurate, even when compared to univariate forecasts for which no use is made of any explanatory variable; see, for example, the discussion in Ashley (1988).

Another solution for the handling of explanatory variables in the forecasting problem is to jointly analyze y_t and X_t within, for example, a vector autoregressive (VAR) model. Although there is evidence of their effectiveness in forecasting key variables, there is a surge in the number of parameters that need to be estimated when the number of variables in X_t increases. The estimation errors can negatively affect the forecast accuracy for the variable of interest, especially in the case of long term forecasting; see, for example, the discussion

in Litterman (1986). In recent years much attention is given to shrinkage methods applied to large sets of explanatory variables including principal components and empirical Bayes methods. In many studies where shrinkage procedures are adopted, convincing evidence of improved forecast accuracy is presented; see also the discussion in Stock and Watson (2012).

2 A new forecasting procedure

In this section we provide the details of our forecasting procedure based on dynamic factor simulations. This development is designed for the long-term forecasting of a key variable by making use of many predictor variables that have non-trivial dynamic structures of their own and have potentially a long-lasting, and typically a cyclical, impact on the key variable. Our particular case of forecasting El Niño events is a clear illustration but such cases also occur in time series related to climate change generally. Although the sea surface temperature is directly associated with the El Niño event, the temperature time series variable is continuous while the El Niño event is a binary event that is defined in terms of a consecutive sequence of temperature exceedances. The La Niña event is defined in a similar way. A complexity arises when we model the sea surface temperature for the purpose of forecasting discrete events which in our case will be El Niño, La Niña and nothing of the two. Apart from the probability forecast, we also develop a simulation-based procedure for computing confidence intervals of the probability forecasts. A final challenge is that the forecast horizon stretches out to 30 months and where forecasts for one year and beyond are of major interest.

Our forecasting procedure consists of three modelling stages: (i) a univariate time series model for key variable of interest, in our case Niño3.4, to produce one-step ahead prediction errors; (ii) a joint analysis of the prediction errors and all potential predictor variables based on a dynamic factor model; (iii) simulation and extrapolation of dynamic factors for the construction of multiple forecasts and confidence intervals for the probability of the El Niño and La Niña events. In this section we provide the details of these three steps and discuss the motivation of the steps in words. More formal motivations and explanations are provided in the Appendix C in which we argue that our simulation-based procedure have better properties for medium-term forecasting when compared to the forecasting based on a dynamic factor model. Some simulation evidence is also provided.

2.1 Step (i): univariate time series analysis

The time series properties of the key variable can be well established in a linear dynamic model. In our study we consider the class of unobserved components (UC) time series of Harvey (1989) but other linear dynamic models can be considered as well including straightforward autoregressive moving average models (ARMA), see Tsay (2010) for an introduction to ARMA models. In Appendix B.1 we provide the details for a time series analysis based on the UC model for a monthly time series y_t that is given by

$$y_t = \mu_t + \gamma_t + \sum_{j=1}^{3} \psi_{j,t} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{1}$$

where μ_t represents the trend component, γ_t represents the monthly seasonal component, $\psi_{j,t}$ represents a stochastic cyclical process with certain amplitude and frequency, for $j = 1, 2, 3$, and ε_t represents the disturbance or noise component. In a linear Gaussian model, we assume that these six unobserved components are generated by linear dynamic stochastic processes depending on Gaussian independent noise terms. In particular, the trend μ_t is modelled as a random walk process, the seasonal γ_t constitutes 11 dummy variables which are modelled as mutual independent random walk processes, the cycles $\psi_{j,t}$, for $j = 1, 2, 3$, are modelled as ARMA processes with low-order lag polynomials but with complex roots imposed on the autoregressive polynomial, and the disturbance component ε_t is assumed to be white noise. The variances of the noise terms and the parameters associated with the cycle components are estimated by the method of maximum likelihood. Once the parameter estimates are given, the unobserved components can be estimated, which is referred to as signal extraction. As part of the estimation procedure, multiple-steps ahead forecasts for y_t can be made, together with standard errors. This methodology relies on the state space approach to time series analysis that has been explored in Durbin and Koopman (2012) and has been considered for the modelling and forecasting of the Niño3.4 time series in Petrova, Koopman, Ballester, and Rodó (2016). In this recent study it is shown that the UC model, extended by a selection of predictor variables, can produce relatively highly accurate forecasts, especially for the medium-term, say from 1.5 to 2.5 years. The high forecast skill is to a large extent explained by the selection of the predictor variables, which are dynamically

relevant to the El Niño evolution specially in the ocean subsurface.

The one-step ahead prediction error of the UC model is defined as

$$v_t^{uc} = y_t - \mathbb{E}_{uc}(y_t|y_1, \ldots, y_{t-1}), \tag{2}$$

where $\mathbb{E}_{uc}$ refers to expectation with respect to the Gaussian density for y_t as implied by the univariate UC model (1) where the unknown parameters are replaced by their maximum likelihood estimates. The predictions errors for the UC model are computed by the Kalman filter which is a linear recursive estimation process for the unobserved components. It implies that, for a given UC model, the prediction error v_t^{uc} is a linear combination of the concurrent and past observations $\{y_1, \ldots, y_t\}$, or in matrix form

$$v^{uc} = L\, y,$$

where $v^{uc} = (v_1^{uc}, \ldots, v_T^{uc})'$, $y = (y_1, \ldots, y_T)'$ and L is a $T \times T$ lower-triangular matrix with unity values on the main diagonal, see Durbin and Koopman (2012, Section 4.13) for a more detailed discussion. The lower triangular elements of matrix $L = L(\psi)$ is a function implied by the UC model and with the parameter vector ψ as its only argument. Since matrix L is invertible, the reverse relationship is also true, that is

$$y = L^{-1} v^{uc}.$$

where L^{-1} is also a lower triangular matrix. The latter relationship is important for the development below since we require the expression that y_t for a given UC model can be reconstructed as

$$y_t = g_t(v_1^{uc}, \ldots, v_t^{uc}), \qquad t = 1, \ldots, T, \tag{3}$$

for a linear function $g_t(\cdot)$ that simply reflects the tth row of matrix L^{-1}.

2.2 Step (ii): dynamic factor model analysis

In the next step we jointly consider the prediction errors of the univariate UC model (for our key variable y_t) and the predictor variables in X_t. The one-step ahead prediction errors v_t^{uc}

are computed by the Kalman filter applied to the UC model. We assume that the number of variables in X_t equals $N-1$. We then formulate the dynamic factor model (DFM) as

$$\begin{pmatrix} v_t^{uc} \\ X_t \end{pmatrix} = \Lambda f_t + \xi_t, \qquad f_t = \Phi_1 f_{t-1} + \ldots + \Phi_p f_{t-p} + \nu_t, \tag{4}$$

for $t = 1, \ldots, T$, where Λ is the $N \times r$ factor loading matrix, f_t is the $r \times 1$ vector with latent dynamic factors, ξ_t is the $N \times 1$ observation disturbance vector, which is assumed to be normally distributed, Φ_i is an $r \times r$ autoregressive coefficient matrix, for $i = 1, \ldots, p$, with p being the order of the vector autoregressive process for f_t, that is VAR(p), and ν_t is the $r \times 1$ normally distributed factor disturbance vector. The r dynamic factors in f_t represent the common dynamic variations in all time series variables in v_t^{uc} and X_t. The dynamic process for f_t is specified as a strictly stationary vector autoregressive process. Hence the underlying assumptions for (4) include the stationarity of v_t^{uc} and all time series variables in X_t. In case of the prediction errors, when the univariate UC model is an adequate model representation for y_t, it is implied that all dynamic information in the time series $y_1, \ldots, y_T$ is accounted for. The prediction errors can therefore be regarded as white noise and certainly stationary. We further assume that all variables in X_t are stationary, possibly after a transformation such as differencing or detrending.

The estimation of the unknown parameters in the matrices Λ, Σ_ξ and Φ_i, for $i = 1, \ldots, p$, is carried by the two-step method of Doz, Giannone, and Reichlin (2011), see Appendix B.2 for the relevant details. In the dynamic factor model as presented in equation (4), the dynamic factors are typically modelled parsimoniously with, say $p = 1$ or $p = 2$, that is a VAR(1) or VAR(2). However this may not be sufficient to capture the persistent and cyclical dynamics in X_t. We therefore introduce an alternative and parsimonious way to introduce X_t and their lags in the forecasting function of y_t.

To fully capture the interactions between v_t^{uc} and X_t, we can focus on the smooth density

$$p_{dfm}(v_t^{uc} | X_1, \ldots, X_T), \tag{5}$$

for $t = 1, \ldots, T$. In case the univariate UC model is the true data generation process for y_t, the prediction errors are white noise and should not be affected by the variables in X_t. We

then simply have $p_{dfm}(v_t^{uc}|X_1,\ldots,X_T) = p_{dfm}(v_t^{uc})$ with $\Lambda = 0$, which implies a white noise process for v_t^{uc}. Any univariate model misspecification due to the omission of the information in $X_1,\ldots,X_T$ will come to light when considering the smooth density (5). Although the DFM is most likely to be misspecified with respect to dynamic specification for X_t, it is of lesser importance when our main focus is on the forecasting of the key variable y_t. We are only interested in a more appropriate model representation for X_t when it increases the forecast accuracy for y_t. We are not seeking the true data generation process for X_t. Hence model (4) is a basic and straightforward model representation for connecting the information in X_t that is relevant for the part of y_t that cannot be explained by its own past.

The simulation smoothing method of Durbin and Koopman (2002) allows the generation of samples from the smooth density (5). Each realised sample for v_t^{uc} is a linear function of $X_1,\ldots,X_T$ while a series of samples from (5) can visualise the amount of variation that is implied by (4) and explained by the collection of predictors in X_t. We denote a simulated prediction residual series from (5) by $v_t^{(i)}$ for $i = 1,\ldots,M$ where M is the number of simulated series that we typically set to $M = 200$.

2.3 Step (iii): forecasting via simulation and estimation

On the basis of the simulated series $v_t^{(i)}$, for $i = 1,\ldots,M$, from the DFM (4), we can generate the corresponding time series y_t on the basis of the UC model (1) since from (3) we have that $y_t = g_t(v_1^{uc},\ldots,v_t^{uc})$. In this way we obtain a sequence of M time series for y_t which we refer to as the set of *ensemble time series* and is denoted and generated by

$$y_t^{(i)} = g_t(v_1^{(i)},\ldots,v_t^{(i)}), \qquad v_t^{(i)} \sim p_{dfm}(v_t^{uc}|X_1,\ldots,X_T), \qquad t = 1,\ldots,T, \tag{6}$$

for $i = 1,\ldots,M$. The ensemble time series $y_t^{(i)}$ is the result of an interaction between the UC and DFM models. This is a key notion of our forecasting method.

The forecast for y_{T+h}, for $h = 1,2,\ldots$, from the univariate UC model can be written as

$$\widehat{y}_{T+h}^{uc} = \mathbb{E}_{uc}(y_{T+h}|y_1,\ldots,y_T) = \mathbb{E}_{uc}(y_{T+h}|v_1^{uc},\ldots,v_T^{uc}), \tag{7}$$

since $v_t^{uc} = g_t(y_1,\ldots,y_t)$ for $t = 1,\ldots,T$. For each ensemble time series (6), we compute the

forecasts as in (7) and we denote these by

$$\widehat{y}_{T+h}^{(i)} = \mathbb{E}_{uc}(y_{T+h}|y_1^{(i)}, \ldots, y_T^{(i)}) = \mathbb{E}_{uc_i}(y_{T+h}|v_1^{(i)}, \ldots, v_T^{(i)}), \qquad h = 1, 2, \ldots, \tag{8}$$

for $i = 1, \ldots, M$, where $\mathbb{E}_{uc_i}$ is expectation with respect to the UC model (1) but with the parameter vector estimated for the ensemble time series $y_1^{(i)}, \ldots, y_T^{(i)}$. The final forecast from our three-step procedure involving dynamic factor simulation (DFS) is obtained by

$$\widehat{y}_{T+h}^{dfs} = \frac{1}{M} \sum_{i=1}^{M} \widehat{y}_{T+h}^{(i)}.$$

This forecast effectively evaluates the forecast function

$$\widehat{y}_{T+h} = \int y_{T+h}(v^{uc}) p_{dfm}(v^{uc}|X) \mathrm{d}v^{uc}.$$

This procedure requires the maximum likelihood estimation of the univariate UC model for the original and M ensemble time series. The DFM model is used to generate ensemble time series to incorporate the predictive contribution of $X_1, \ldots, X_T$ for the key variable y_t. The actual forecasting of the X_t variables is not needed for the forecasting of y_t.

2.4 Discussion and review of the new forecasting procedure

The incorporation of predictor variables in the UC model (1) is straightforward since we can simply add a regression term such as $X_t\delta$, where δ is the $(N-1) \times 1$ vector of regression coefficients, to the right-handside of (1). The estimation of δ, together with the parameter vector ψ, can be carried out within the state space approach of time series analysis. However, in cases where we have many variables in X_t and the dependence of y_t on X_t is strong and lasts over many lags of X_t. The number of parameters to estimate increases rapidly and it is well known that heavy parameterised models are typically not superior in producing accurate forecasts, and especially not for longer-term forecasts. Variable selection and lag-selections procedures are hazardous without good knowledge of the strong dependencies amongst the variables and their intricate and strong dynamic inter-linkages. Finally, in the current context in which the UC model with X_t's will also be used for the forecasting of

y_t, it clearly also requires forecasts for the variables in X_t. The effort to provide accurate forecasts at each time a forecast of y_t is required can be regarded as a severe task and will not necessarily lead to a good forecast for y_t.

To circumvent a heavy parameterised UC model but still allowing predictor variables (including their lags) to contribute in the forecasting of the key variable, various methodologies have been developed. For example, the use of principal components within a dynamic factor analysis have been heavily explored in the macroeconomic forecasting literature with key contributions by Stock and Watson (2002) and Doz et al. (2011). Although the use of the principal components addresses the challenge of variable selection, it does not address the lag-selections for each variable. To capture the strong serial dependencies amongst the variables y_t and those in X_t in a dynamic factor analysis, we still require the empirical identification of the dynamic inter-linkages between the key variable y_t and the selected dynamic factors as well as the dynamic structures within the dynamic factors. Hence, we only have partially addressed the heavy parameterisation issue. We therefore have developed the DFS solution to address these issues by relying both on a univariate time series analysis, for example based on the UC model, and on dynamic factor analysis. This simulation method is simple and also provides a general solution for the problem of forecasting non-linear signals.

Our proposed procedure circumvent the need to forecast X_t's completely while the explanatory variables still play an important role in the forecasting of y_t. It is a parsimonious procedure and it is based on the reconstruction of a time series y_t that takes into account the information of explanatory variables. The construction of these ensemble time series is computationally fast since we adopt the simulation smoother of Durbin and Koopman (2002). We emphasise that the simulated prediction errors $\tilde{v}_1^{(i)}, \ldots, \tilde{v}_T^{(i)}$ from the DFM are by construction functions of the explanatory variables $X_1, \ldots, X_T$, these are their lag, concurrent and lead values for y_t with $1 \leq t \leq T$. The univariate UC model of the initial analysis is revisited for each ensemble time series to re-estimate the parameters and to produce forecasts for the ensemble $y_{T+h}^{(i)}$, for $h = 1, 2, \ldots$. Our final forecast is simply the average of these M forecasts.

The procedure is rather general and can be implemented using different choices of models and estimation methods. For example, a simple autoregressive model can be considered in the first step while the second step can be based on a few explanatory variables and one

principal component. The dynamic factor model in the second step is adopted as a data reduction technique to handle potentially many explanatory variables and as a framework to generate paths of the prediction errors conditional on past and concurrent observations of the explanatory variables. Other shrinkage methods can be considered but we require a model representation to enable the simulation from $p(v_t|X_1, \ldots, X_T)$. The prediction errors are a key ingredient in the procedure as they are constructed with the attempt to remove the serial dependence in the time series variable of interest. Hence we separate the dependence of y_t from its own past (serial correlation) and the dependence of y_t on the information present in the current and past observations of the explanatory variables. We first adopt an appropriate univariate time series and obtain the prediction errors and, subsequently, we jointly model the prediction errors and the explanatory variables within a dynamic factor model. The next step of reconstructing prediction errors conditional on, or as a function of, $X_1, \ldots, X_T$, is an efficient way to incorporate this information into the forecast for $\widehat{y}_{T+h}$. A more formal justification of this method is presented in Appendix C. Our proposed forecasting method can also be motivated from a model misspecification argument since it provides the ingredients to carry out a Durbin-Wu-Hausman test as formulated by Hausman (1978) with respect to the forecast for $\widehat{y}_{T+h}$. It can be shown that the forecasts obtained in the first and third steps are consistent while in the third step the forecasts are more efficient. Since we adopt the dynamic factor model in the second step, we can also carry out a test whether the X_t's are endogenous.

3 Empirical in-sample results for the case of El Niño

3.1 The data

Our data set includes the monthly time series of temperature values which is referred to as the Niño3.4 time series and which is defined as the area-averaged sea surface temperature in the region (5 ° N - 5 ° S, 170 ° W - 120 ° W). In this area the El Niño events are identified, see also the discussion in Barnston, Chelliah, and Goldenberg (1997). The National Centers for Environmental Information (NOAA) defines an El Niño or La Niña event as a phenomenon in the equatorial Pacific Ocean characterised by a five consecutive 3-month running mean

of sea surface temperature (SST) anomalies in the Niño 3.4 region that is above (below) the threshold of +0.5°C (-0.5°C). This standard of measure is known as the Oceanic Niño Index (ONI). The ONI index is calculated as the average sea surface temperature in the Niño 3.4 region for each month, and then averaged with values from the previous and following months. This running three-month average is compared to a 30-year average. The observed difference from the average temperature in that region, whether warmer or cooler, is the ONI value for that 3-month "season". This definition is obtained from the website of NOAA, `https://www.ncdc.noaa.gov/`.

In our empirical study we consider the Niño3.4 time series as the variable of key interest and denote it as y_t. This time series is observed from January 1982 to the end of 2015, that is 34 years of data and in total 408 monthly observations. For this period, we also have all observations for 24 explanatory variables which consist of physical measures of zonal wind stress and sea temperatures at different depths in the ocean and at different locations. These variables are selected because they are relevant to the dynamical processes for the generation and evolution of El Niño events; see Petrova et al. (2016) for further details. We do have observations available for beyond 2015 but only for a selection of variables, not for all of them.

3.2 Some computational details

In the implementation of our new forecasting procedure, all steps rely on the linear Gaussian state space model as discussed in Durbin and Koopman (2012). In the first step, we consider the unobserved components time series model (1) with stochastic components consisting of a level specified as a random walk, a time-varying dummy seasonal, three stochastically time-varying cycles, and an irregular specified as a white noise process. The parameters are estimated by the method of maximum likelihood by means of the numerical maximisation of the loglikelihood function that is evaluated by the Kalman filter. The same Kalman filter is also used for evaluating the in-sample prediction errors v_t^{uc} and for computing the forecasts for y_{T+h} for $h = 1, 2, \ldots$. In the second step, the dynamic factor model is formulated in state space form with the system matrices obtained from the Doz et al. (2011) two-step procedure based on principal components. The number of principal components (equals

the number of factors) is equal to what number is necessary to capture 95% of the sample variation in $(v_t^{uc}, X_t')'$. The dynamic factors are specified as the VAR(p) process with $p = 2$. Once the dynamic factor model is presented in state space form, the simulated values $\tilde{v}_t^{(i)}$ from $p_{dfm}(v_t^{uc}|X_1, \ldots, X_T)$ are obtained after a slight modification of the simulation smoother of Durbin and Koopman (2002). In the third step, we repeat the univariate forecast procedure from step 1 for all ensemble time series for which the parameters are re-estimated by maximum likelihood for each series. Since the dynamic properties of the ensemble time series are very similar, good starting values are available and the repeated estimations do not take much computer time. The sample means of the forecasts for each horizon h are recorded as our final forecasts. We will provide empirical evidence that these forecasts outperform the forecasts from a range of relevant and competitive benchmark models.

3.3 Descriptive data analysis

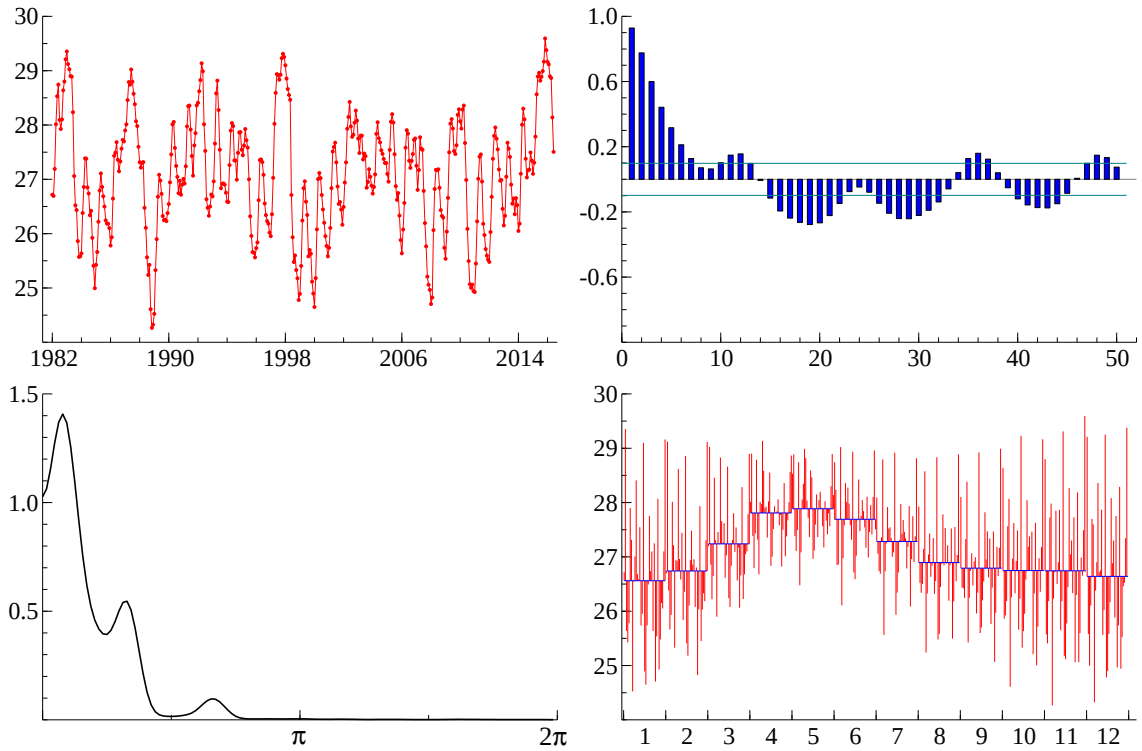
To gain some initial insights into the dynamic properties of the Niño3.4 time series, we present in Figure 1 the time series, its sample autocorrelation function, the sample spectrum (computed from the sample periodogram and subject to some smoothing constant), and the yearly time series plots for each month and the corresponding averages over all 34 years. We may conclude that the time series has stationary features but is highly persistent and cyclical structures appear to be present in the dynamics of the time series. In particular, we clearly observe at least three local peaks in the sample spectrum. This may indicate that three cyclical processes with different periodicities are important ingredients in the data generation process of the Niño3.4 time series. We may also conclude that the time series contains clear seasonal variation in the mean. It is interesting to notice that the seasonal variation in the lowest and highest percentiles are both different from the mean and from each other.

3.4 Decomposition of the Niño3.4 time series

As our univariate time series model, we adopt the unobserved components time series model as given by (1) and discussed in detail in Appendix B.1. The initial findings above suggest that we may decompose the Niño3.4 time series variable into stochastic processes for trend,

Figure 1

Niño3.4 time series and its dynamic properties



The figure shows the Niño3.4 time series and some diagnostic graphs. In clock-wise order: monthly time series plot y_t from 1982 until 2015, the sample autocorrelation function of y_t, the yearly plots of y_t for each month and with its periodic average; the sample estimate of the spectrum of y_t (smoothed version of the sample periodogram).

seasonal, three cycles, and irregular components. The parameters of the decomposition model are estimated by maximum likelihood as this is discussed in Appendix B.1. The different time series components are extracted from the time series using the Kalman filter and smoothing methods, see Durbin and Koopman (2012).

The maximum likelihood estimate of the variance σ_η^2 is numerically very close to zero so that the stochastic trend effectively reduces to a fixed intercept. It indicates that the Niño3.4 time series does not have nonstationary properties (weakly stationary). Similarly, the seasonal component γ_t also has reduced to a set of 11 fixed monthly seasonal dummy effects during the estimation process. Also the estimate of variance σ_ε^2 for the disturbance term is very close zero. The estimated variances of the three cyclical processes are all non-zero, although the variance for the second cycle is relatively close to zero. Hence, after the estimation procedure, the decomposition model for the Niño3.4 time series has

reduced to a linear regression model with an intercept, 11 fixed monthly seasonal dummy effects and autoregressive moving average (ARMA) disturbances for which the autoregressive polynomial has a multiple of complex roots.

For each of the three stochastic cycle components in the decomposition model we have estimated parameters for persistence, frequency and variance. Their values for persistence are 0.96, 0.99 and 0.98 while the estimates for frequency are 1.45, 2.46 and 4.44, respectively. These correspond approximately to the three local modes of the sample spectrum for the Niño3.4 time series. In Figure 2 we present all estimated components, including the estimated intercept and seasonal dummies, but in a time series plot. The sum of these estimated components equals the original time series y_t. The third cycle component with the lowest frequency appears to account for most of the variation in the time series since it has the highest amplitude. The first cycle with the highest frequency can be associated with the short-term cyclical dynamics which appears to account for more variation than the seasonal component. The second cycle with the intermediate frequency is smooth and accounts for some bi-annual systematic variation in the time series but its amplitude is relatively small compared to the seasonal and other cyclical components.

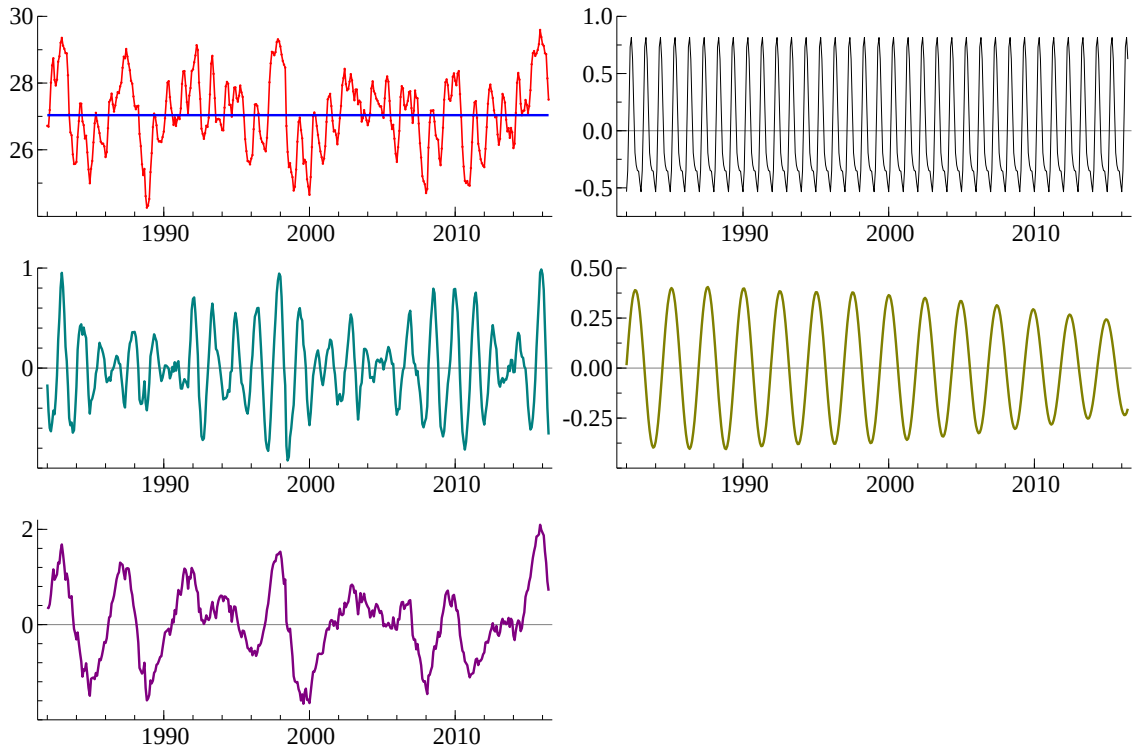
The standardised one-step ahead prediction errors should be white noise and normally distributed with mean zero and unity variance, when the model is well-specified. We can verify these assumptions through a range of diagnostic tests. In Figure 3 we provide an insight of the residual properties by presenting a time series plot of the standardised residuals together with the correlogram, spectrum and histogram. These diagnostic plots confirm the white noise properties of the errors to some extent although some weak cyclical behaviour appears to be unaccounted for. From the histogram we learn that the Gaussian distributional assumption does not appear to be seriously violated despite the outlying value at the beginning of 1998.

3.5 The ensemble residuals

To provide some more insights into the role of the ensemble time series in the third step of our forecasting procedure, we present some features of the ensemble residuals. The simulated one-step ahead prediction errors from the smooth density $p_{dfm}(v_t^{uc}|X_1,\ldots,X_T)$ are referred

Figure 2

Decomposition of Niño3.4 time series

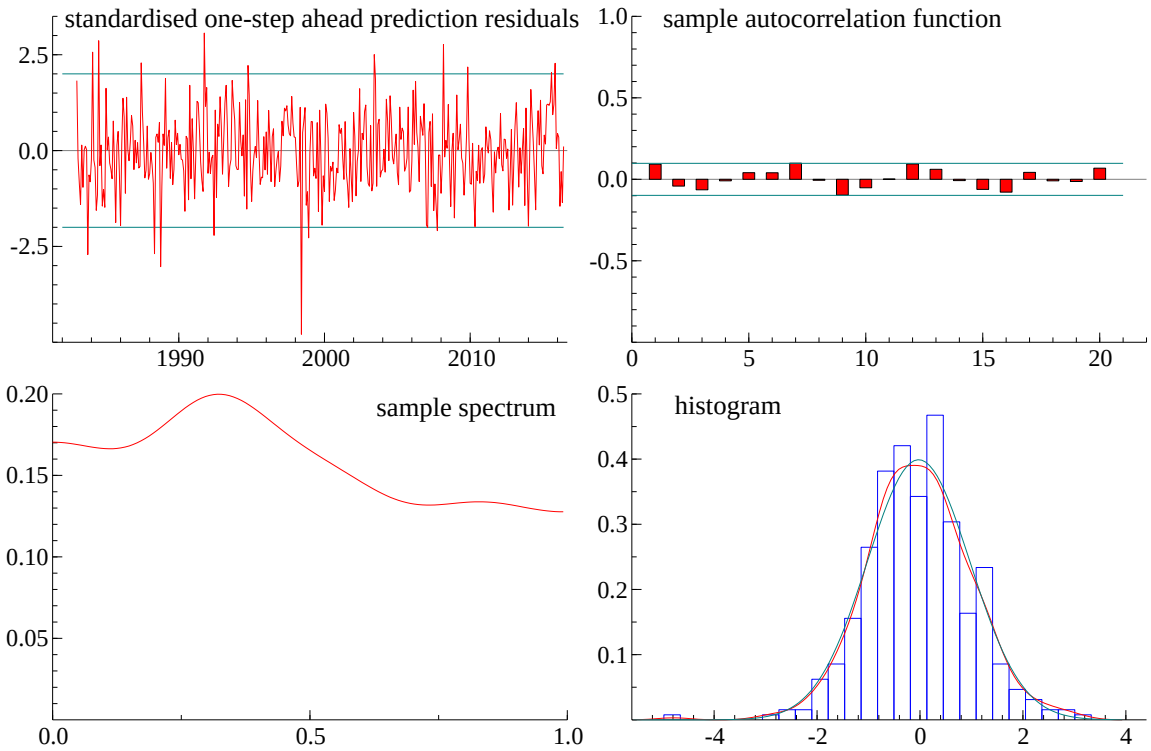


The figure presents the estimated components from the univariate decomposition model applied to the Niño3.4 time series. In the first row, from left to right, we present the estimated trend component μ_t, together with the Niño3.4 time series, and the estimated monthly seasonal component γ_t, respectively. The variances for both of these components are estimated to be zero and hence the components have reduced to fixed effects. In the other rows, the three estimated cycle components are presented. The disturbance variance is estimated to be zero and hence the disturbance component has vanished from the model.

to as ensemble residuals and they are generated by the simulation smoother of Durbin and Koopman (2002) in the third step of the procedure. In Figure 4 we present a sequence of 100 ensemble prediction error time series, together with the original one-step ahead prediction errors v_t^{uc}. It is interesting to see that this set of ensemble residuals form a band that goes through the one-step ahead prediction errors, despite the fact that unconditionally the residuals can be regarded as white noise. The ensemble residuals can be considered as potential signals that are relevant for v_t^{uc} and are constructed from the information in $X_1, \ldots, X_T$. The main path of the ensemble residuals reflects the signal while the band-width reflects the amount of variation implied by the X_t's. To illustrate the significant contribution of the X_t's on the prediction residuals v_t^{uc}, we compute for each ensemble series the F-test for the fit in the regression model with v_t^{uc} as the dependent variable and the ith ensemble

Figure 3

Diagnostic graphics for decomposition Niño3.4 time series



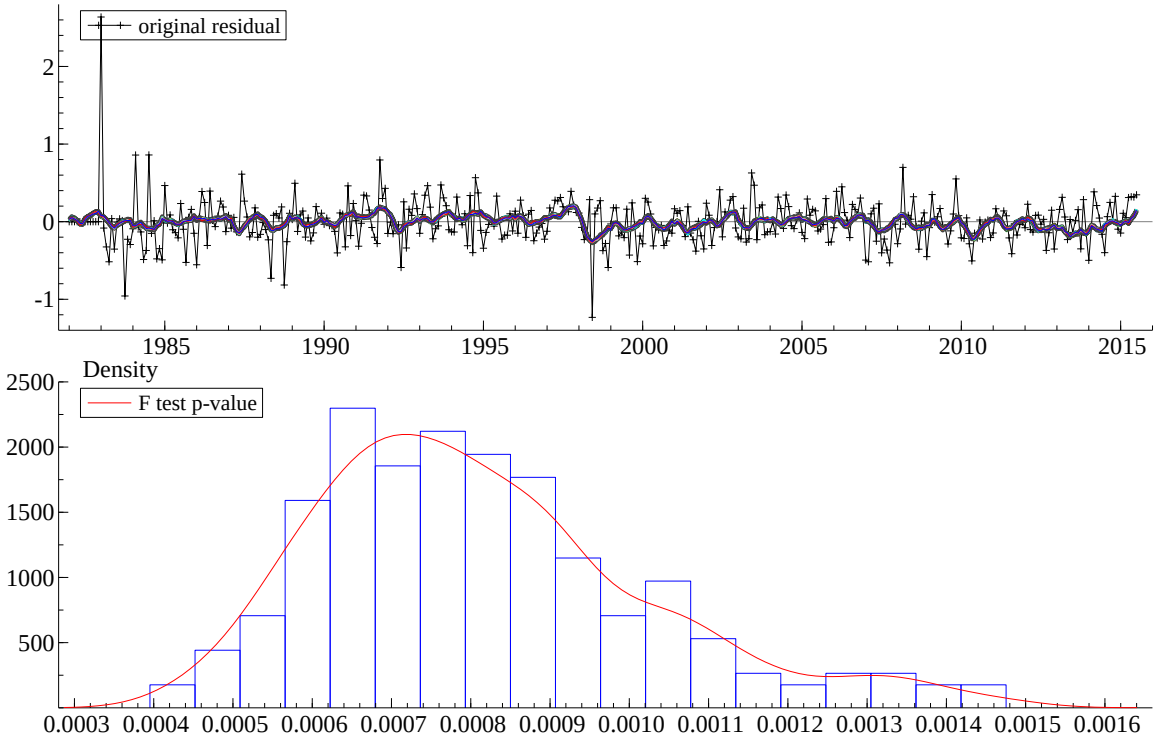
The figure presents four diagnostic graphs for the univariate time series analysis based on the UC model (1) applied to the Niño3.4 time series. In the first row, from left to right, we present the standardised one-step ahead prediction residuals and its sample autocorrelation function (upto 20 lags). In the second row, from left to right, we present the corresponding sample spectrum (smooth estimate of the sample periodogram) and the histogram of the standardised prediction errors.

residual $v_t^{(i)}$ as the explanatory variable. We record the corresponding p-value of the F-test and present the histogram of the 100 p-values in Figure 4. It is clear that each individual ensemble residual series contribute significantly in the fit of the prediction residuals v_t^{uc} with all p-values smaller than 0.002. However, it also shows that some ensembles produce a better fit than others.

To investigate the individual role of a variable in X_t and to further illustrate the rich information that is contributed by X_t to the ensemble residuals, we present in Figure 5 the average correlation of $\tilde{v}_t^{(i)}$ with a selection of variables in X_t and their lagged values, up to lag 30. The correlations are computed for each ensemble residual series $\tilde{v}_t^{(i)}$ and then the average is taken over $i = 1, \ldots, M$. This procedure is similar in design and has the same aim as the R^2 plots of Stock and Watson (2002, see their Figure 1) between principal components and individual explanatory variables. Some of the panels of Figure 5 clearly show a seasonal

Figure 4

Prediction and Ensemble residuals for Niño3.4 time series



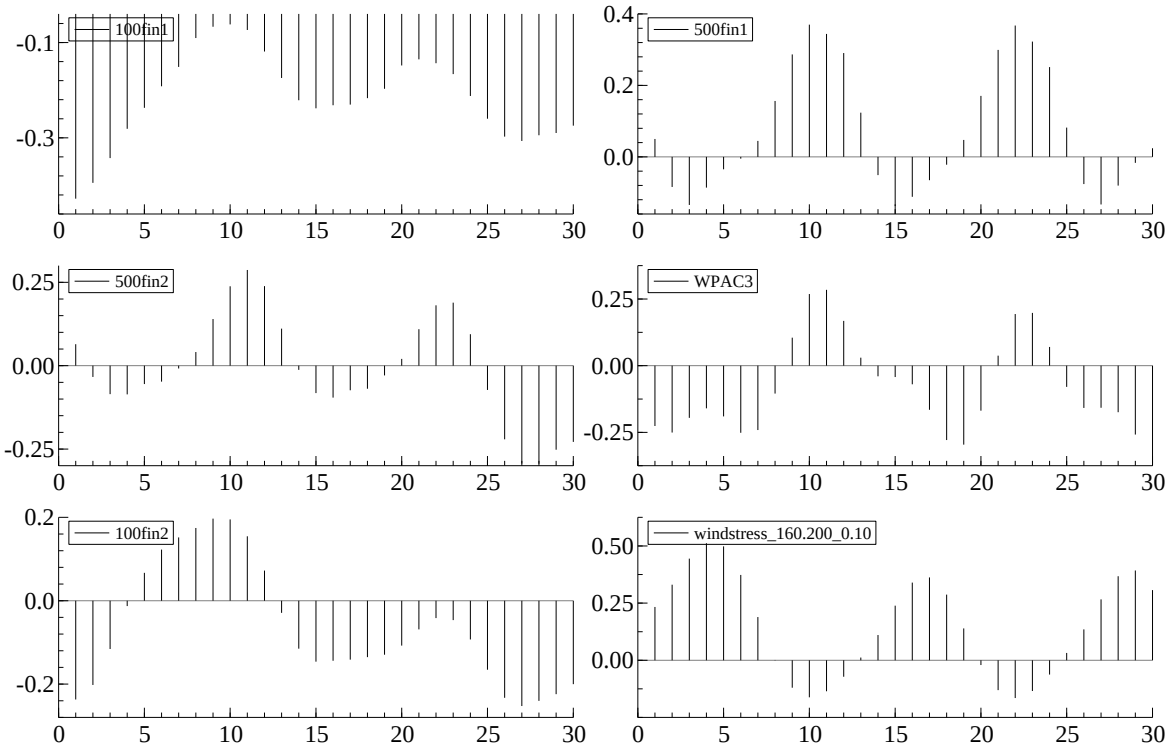
The figure presents the one-step ahead prediction errors from the UC model and obtained as part of the first step of our forecasting procedure. The prediction errors are computed by the Kalman filter applied to the unobserved components time series model with trend, seasonal, and three cycles for the Niño3.4 time series. In addition, $M = 200$ ensemble prediction errors sampled from the conditional density $p_{dfm}(v_t^{uc}|X_1, \ldots, X_T)$ implied by the second step dynamic factor model are plotted in the top panel. They appear as a band through the one-step ahead prediction errors which is obtained through $X_1, \ldots, X_T$. The bottom panel shows the histogram of p-values of the F-test for M regressions $v_t^{uc} = c + \beta \tilde{v}_t^{(i)} + \epsilon_t$. The p-values are very low indicating that the estimated parameters are highly significant which means that a significant part of the variation in v_t^{uc} is explained by the simulated signal $\tilde{v}_t^{(i)}$.

pattern with roughly one year period. Hence it is evident that the seasonality in the Niño3.4 time series covariates with the seasonality of some of the explanatory variables as expected, and thus X_t helps to capture this seasonal variability better during forecasting. On the other hand, we are mainly interested in explaining those parts of v_t^{uc} that are not seasonal, as El Niño is an oscillation on inter-annual time scales. Therefore, we are especially interested in the high correlations between X_t and v_t^{uc} in Figure 5, those that do not correspond with seasonal periodicities. For example, in the case of the "100fin2" variable we observe that it continuously explains variability for the lags 10-30, which is expected as typically before El Niño during this time there is a subsurface anomalous warming of the ocean in the area where this variable is defined, which warming later propagates and plays an important role

in the generation of El Niño in the eastern Pacific. Similarly, in the case of the "500fin2" variable there is a high correlation for the lags 16-30, again corresponding to this early subsurface warming of the ocean at greater depths. Finally, in the case of the wind stress variable we notice that there is a continuous high correlation between lags 0-16, which corresponds to a time before El Niño when typically westerly wind anomalies occur in the central Pacific and assist in the development of the phenomenon. All of these variables were originally constructed with the purpose to capture important dynamical information about El Niño at its very early development stages and corresponding to very long lead times. With our procedure we aim to extract every bit of this information, and use it efficiently and parsimoniously in our forecasting of the El Niño events.

Figure 5

The correlations between the ensemble residual series and X_t



The figure presents the average correlations between $\tilde{v}_t^{(i)}$ and a selection of variables in X_t, contemporaneously and for its lags from 1 to 30. Some seasonal patterns can be detected but also it is shown that some variables show a high correlation at certain lags but not at others. The selected explanatory variables are indicated in the graphs. We have selected 6 out of 24 explanatory variables.

4 Forecasting Niño3.4 time series and El Niño events

4.1 Design of forecasting study

In the first part of our empirical forecasting study we consider the time series Niño3.4 as the variable of interest y_t that we want to forecast h-steps ahead for $h = 1, 2, \ldots, 30$. Hence the maximum forecast window is 2.5 years. The time series is observed from January 1982 to the end of 2015, that is 34 years of data and in total 408 monthly observations. For this period, we also have all observations for the 24 explanatory variables; see Petrova et al. (2016) for more details and see also the discussion in the previous section. We collect all 24 variables in the vector of explanatory variables X_t. For producing the forecasts of y_t we adopt different forecast modelling approaches and methods, including our three-step forecasting procedure, indicated as DFS. To evaluate the quality of the different forecast strategies we carry out a rolling-window forecasting study. We consider an estimation window of 275 monthly observations. The first window starts in May 1982 and we use the observations in the estimation window to carry out the forecast procedure to obtain $\hat{y}_{T+h|T}$ for $h = 1, 2, \ldots, 30$. The second window starts in June 1982 and we use again the next 275 observations to produce the next 30 forecasts by our procedure (all estimation tasks are repeated). In this setting we can repeat this 99 times and hence obtain 99 out-of-sample forecast errors for each horizon $h = 1, 2, \ldots, 30$. We then compute measures of forecast precisions such as the root mean squared error (RMSE), based on these 99 forecast errors, for each h.

The results for our DFS method are the focus of our presentations. We compare different measures of predictive accuracy of the DFS method with those obtained from alternative model-based forecasting methods. We have considered the following models. (i) The seasonal autoregressive moving average (SARMA) model is specified with seasonal autoregressive and moving average lag polynomials of length 12 and with the corresponding regular polynomials of length 2 and 1, respectively. This specification is based on the Bayesian information criterion (BIC) for which these polynomial lengths have shown to have the smallest in-sample BIC values. After parameter estimation by maximum likelihood, the model is represented in state space form such that the Kalman filter is used to produce all multi-step ahead forecasts. (ii) The unobserved components (UC) time series model or UCM as specified in the first step of our DFS forecasting procedure. (iii) The VAR forecasts are based on an unrestricted

VAR(2) estimation for all 25 variables in the observation vector $(y_t, X_t')'$. The lag length of 2 is also determined by the smallest in-sample BIC value. After estimation via regression, the VAR(2) model is formulated in state space form such that the Kalman filter is used to produce all multi-step ahead forecasts. (iv) The forecasts based directly on the principal components used in the second step of DFS are computed as for the Stock and Watson (2002) approach. The number of principal components that are used is determined by having 95% explained in the data matrix for X_t; this is indicated by the eigenvalues of the sample variance matrix. We have typically used 5 to 7 principal components as a result. The inclusion of lags for y_t and for the principal components in the Stock and Watson procedure for forecasting is determined by in-sample BIC. (v) The dynamic factor model (DFM) is the basic version with the observation vector $(y_t, X_t')'$ being linearly dependent on a set of five independent dynamic factors, each modelled as a stochastic cycle, see the discussion in Appendix B.2 and Durbin and Koopman (2012, Chapter 3). The model is cast directly in state space form and the loading, persistence and variance matrices are estimated by numerically maximizing the loglikelihood function that is obtained from Kalman recursions. (vi) The least absolute shrinkage and selection operator (LASSO) is used to select predictors from the collection $(X_t', X_{t-1}', \ldots, X_{t-36}')'$ which is 888-dimensional with the lag of 36 suggested by Petrova et al. (2016). The LASSO threshold or shrinkage parameter is chosen by BIC and the typical number of nonzero coefficients is 25. (vii) The collapsed dynamic factor model (CDFM) by Bräuning and Koopman (2014) summarizes information from the 889-dimensional vector $(y_t, X_t', X_{t-1}', \ldots, X_{t-36}')'$ using principal components and form a DFM with y_t and extracted principal components. Based on the IC_p$_1$ criterion of Bai and Ng (2002), the typical number of principal components is 30, close to the number of predictors selected by LASSO. For our DFS forecasting procedure and the above-mentioned model-based forecasting methods, we report the RMSE, MFLE and MRPS defined in the next section for each forecast horizon $h = 1, 2, \ldots, 30$ with the Diebold and Mariano (1995) test to verify the superior predictive accuracy of the forecasts from our proposed method. Finally, we carry out a forecast exercise for the probability of El Niño events using the DFS method.

4.2 Forecasting loss functions and precision criteria

We measure the precisions of out-of-sample forecasting using different loss functions and from different perspective. The forecasting exercises are based on multiple steps ahead forecasting in a rolling window setting. In this way we obtain a specific number of forecast errors for each method and for each forecast window length. We indicate a particular rolling window by i and a forecasting length by h.

To measure the predictive accuracy, we define the following loss differential function for a rolling window i and a forecast length h as

$$d_{i,h} = L(e_{i,h}^{(j)}) - L(e_{i,h}^{(k)}),$$

for some loss function $L(\cdot)$ and h-step ahead forecast errors $e_{i,h}^{(j)}$ and $e_{i,h}^{(k)}$ obtained from models j and k, respectively. In our forecasting design, we have $M = 99$ rolling windows with each containing $T = 275$ data points, Harvey, Leybourne, and Newbold (1997) suggests the following Diebold-Mariano (DM) test statistic corrected for small sample size

$$DM_h^* = \sqrt{\frac{M + 1 + h^2 - 3h}{M}} DM_h \sim \text{Student's } t_{M-1}$$

where DM_h is the Diebold and Mariano (1995) test statistic for the h-step ahead forecast loss differential as given by

$$DM_h = \bar{d}_h / \sqrt{\frac{1}{M}\hat{\sigma}_{T,M}^2},$$

with average loss differential $\bar{d}_h = \frac{1}{M}\sum_{i=1}^{M} d_{i,h}$ and where the heteroskedastic and autocorrelation (HAC)-consistent variance estimator of $\text{Var}(\sqrt{M}\bar{d}_{i,h})$ is given by

$$\hat{\sigma}_{T,M}^2 = \hat{\gamma}(0) + 2\sum_{p=1}^{P} \alpha_p \hat{\gamma}(p),$$

see Giacomini and White (2006) for a discussion, with the Bartlett weights $\alpha_p = 1 - p/P$ with P being the integer part of $M^{1/4}$, for $p = 1, \ldots, P$. When we adopt the quadratic spectral weights, the values for our forecast precision measures have only changed marginally in our study.

Tables 1 and 2 present the root mean squared forecast error (RMSE) of model j as given by

$$RMSE_h^{(j)} = \sqrt{\frac{1}{M}\sum_{i=1}^{M} L(e_{i,h}^{(j)})}, \qquad h = 1, ..., 30,$$

which summarizes the predictive accuracy based on the squared error loss function $L(e_{i,h}^{(j)}) = e_{i,h}^{(j)^2}$ which we use in the HLN_DM test. Table 3 and 4 shows the h-step ahead mean linex forecast error $MLFE_h^{(j)}$ of model j defined as

$$MLFE_h^{(j)} = \frac{1}{M}\sum_{i=1}^{M} L(e_{i,h}^{(j)}), \qquad h = 1, ..., 30, \tag{9}$$

which summarizes the predictive accuracy based on the linex loss function $L(e_{i,h}^{(j)}) = \exp(\beta e_{i,h}^{(j)}) - \beta e_{i,h}^{(j)} - 1$; this loss function has been originally proposed by Varian (1975). The linex loss function is a combined measure of loss in point forecast and loss in the forecasting direction of change. Given the asymmetry of the linex loss function, it is widely used in financial risk management where underestimating *value-at-risk* (VaR) is more costly than overestimating it. The parameter β measures the aversion towards either negative ($\beta > 0$) or positive ($\beta < 0$) forecast errors. We choose $\beta = 1$ since underestimating the Niño3.4 forecast increases the probability of missing an El Niño event.

Table 5 shows mean ranked probability score defined in the same way as (9) to summarise the loss from some chosen models in forecasting the events of El Niño and La Niña. The rank probability score (RPS) statistic, Epstein (1969), is given by

$$RPS = \frac{1}{2}\sum_{k=1}^{3}\left(\sum_{j=1}^{k}(\hat{p}_{j,i,h} - e_{j,i,h})\right)^2,$$

where $j = 1$ indicates an El Niño event, $j = 3$ a La Niña event, and $j = 2$ corresponds to no event. The probability $\hat{p}_{j,i,h}$ is the forecasted probability for each of the three categories $j = 1, 2, 3$ for window i and lag h. The binary variable $e_{j,i,h}$ is the actual outcome (0,1) for each of the three categories for window i and lag h. This scoring rule is sensitive to distance by taking into account the probability mass assigned to all categories and not only the category of the observed outcome as for example in a log loss function. We average

the RPS over all forecasts (for a specific forecast horizon h) from the rolling windows. This average RPS provides a good indicator of the quality of our event forecasts. The probabilities $\hat{p}_{j,i,h}$ are obtained by simulation. For each of the simulated series $\nu_t^{(i)}$ obtained in Step (ii) of our proposed method, we draw $N = 100$ series with the simulation smoother. We extend the series $\nu_t^{(i)}$ with 30 missing values to obtain a cloud of $h = 1, \ldots, 30$ step ahead forecasts. Once we simulate a cloud of forecasts we count the number of times an El Niño event, La Niña event, or no event occurs within a simulate 30-step ahead forecast. The simulated probabilities $\hat{p}_{j,i,h}$ are the number of times each event occurs divided by the total number of draws. The confidence interval for each probability can be determined by

$$\hat{p}_{j,i,h} - Z \times \sqrt{\hat{p}_{j,i,h}(1 - \hat{p}_{j,i,h})/N}, \qquad \hat{p}_{j,i,h} + Z \times \sqrt{\hat{p}_{j,i,h}(1 - \hat{p}_{j,i,h})/N},$$

with Z the value corresponding to the desired interval. We refer to Figures 9, 10, and 11 for the visualisation of simulated probabilities for selected El Niño events.

4.3 Forecasting results for the 3.4Niño time series

Let us first briefly discuss the structure and content of the Tables 1 to 5. A cell with only a value and without any symbol (except for the column with header UC_DFS) implies that our DFS method outperforms the other model at a 5% level of significance. A shaded cell indicates that our DFS method underperforms another model at the 5% level. Single asterisk "$*$" indicates that the DFS method outperforms the other model at a significance level of only 10% level. Double asterisks "$**$" indicates no significant difference in predictive accuracy between US_DFS and another model.

Our DFS method is categorised as a multivariate forecasting method in Table 1 for the apparent reason that it utilises relevant variables collected in X_t. In terms of squared forecast errors, which is a measure for point forecast accuracy, DFS significantly outperforms all other multivariate models from 5-step ahead on, except for LASSO and VAR models which are occasionally inferior to DFS only at 10% level, such as for $h = 7$, 8, 9 and 29. The VAR model does well in the short horizon, surpassing the DFS method at $h = 1$ and 3, and especially for $h = 1$, the RMSE of VAR is only 0.2422, the lowest among all multivariate models considered and 33% lower than that of DFS which is 0.3626. A reason for this is that

Table 1

RMSE of Multivariate Models

The table reports the root mean squared forecast error $RMSE_h$ for multivariate benchmark models as explained in Section 4.1. A cell without a symbol (except for the column with header UC_DFS) means the DFS method outperforms another model at the 5% level. A shaded cell indicates that the UC_DFS underperforms compared to another model at the 5% level. Single asterisk "$*$" indicates that the UC_DFS model outperforms another model only at the 10% level. Double asterisks "$**$" indicates no significant difference in predictive accuracy between US_DFS and another model.

Step h	DFM	CDFM	LASSO	S&W	VAR	UC_DFS
1	0.4857*	0.3749**	0.4792	0.4966	0.2422	0.3626
2	0.7953*	0.6059*	0.6255*	0.7378	0.4576**	0.4886
3	1.0339	0.7834	0.7534	0.9518	0.5890	0.5976
4	1.1823	0.8632	0.8365	1.1203	0.6626**	0.6879
5	1.2718	0.8909	0.8887**	1.2519	0.7567**	0.7505
6	1.3163	0.9258*	0.9230	1.3582	0.8718	0.7875
7	1.3297	0.9798*	0.9278*	1.4465	0.9780*	0.8119
8	1.3150	1.0121	0.9509*	1.5217	1.0548*	0.8276
9	1.2634	1.0012	0.9945*	1.5785	1.1162	0.8299
10	1.1765	0.9870	1.0130	1.6205	1.1566	0.8277
11	1.0824	0.9973	1.0137	1.6618	1.1929	0.8227
12	1.0273	1.0280	1.0117	1.7329	1.2012	0.8194
13	1.0253	1.0460	1.0276	1.8204	1.1648	0.8189
14	1.0522	1.0407	1.0246	1.8978	1.1050	0.8271
15	1.0833	1.0414	0.9897	1.9583	1.0640	0.8131
16	1.0996	1.0512	0.9954*	1.9943	1.0547	0.8143
17	1.1020	1.0490	1.0133	2.0020	1.0488	0.8122
18	1.0945	1.0277	1.0412	1.9867	1.0414	0.8075
19	1.0813	1.0004	1.0219	1.9617	1.0408	0.7994
20	1.0644	0.9845	0.9939	1.9235	1.0446	0.7886
21	1.0446	0.9878	0.9412*	1.8520	1.0352	0.7808
22	1.0235	1.0059	0.9538	1.7594	1.0200	0.7729
23	1.0067	1.0356	0.9724	1.6849	0.9930	0.7637
24	0.9996	1.0603	0.9497	1.6642	0.9654	0.7565
25	0.9990	1.0621	0.9684	1.7092	0.9403	0.7518
26	1.0015	1.0444	0.9578	1.7853	0.9214	0.7766
27	1.0132	1.0370	0.9418	1.8525	0.9192*	0.7665
28	1.0295	1.0469	0.9383	1.8959	0.9282	0.7819
29	1.0490	1.0608	0.9326*	1.9242	0.9384*	0.8063
30	1.0653	1.0568	0.9419	1.9577	0.9484	0.8304

Table 2

RMSE of Univariate Models

The table reports the root mean squared forecast error $RMSE_h$ for univariate benchmark models as explained in Section 4.1. A cell without a symbol (except for the column with header UC_DFS) means the DFS method outperforms another model at the 5% level. A shaded cell indicates that the UC_DFS underperforms compared to another model at the 5% level. Single asterisk "$*$" indicates that the UC_DFS model outperforms another model only at the 10% level. Double asterisks "$**$" indicates no significant difference in predictive accuracy between US_DFS and another model.

Step h	AR(6)	ARMA(4,3)	LL	SARMA(12,2,1)	UCM
1	0.3089	0.3145	0.4343	0.2954	0.2349
2	0.6014**	0.6071*	0.7874	0.5574	0.3956
3	0.8281	0.8347	1.0624	0.7584	0.5251**
4	0.9630	0.9681	1.2565	0.8893	0.6413
5	1.0291	1.0328	1.3902	0.9673	0.7417**
6	1.0455	1.0482	1.4755	1.0022	0.8068*
7	1.0403	1.0457*	1.5264	1.0166	0.8474
8	1.0330	1.0429	1.5451	1.0262	0.8749
9	1.0280*	1.0411	1.5174	1.0296	0.8916
10	1.0197	1.0290	1.4448	1.0226	0.8961
11	1.0164	1.0117	1.3592	1.0131	0.8938
12	1.0235	1.0043*	1.3175	1.0110*	0.8940*
13	1.0247*	1.0029*	1.3471	1.0037*	0.8975*
14	1.0100	0.9988	1.4254	1.0025	0.8981
15	0.9921	0.9964	1.5100	1.0106	0.8973
16	0.9867	1.0008	1.5697	1.0186	0.8963
17	0.9882	1.0045	1.5996	1.0202	0.8939*
18	0.9887	1.0018	1.6057	1.0137	0.8896*
19	0.9878	0.9968	1.5971	1.0062	0.8833
20	0.9872	0.9932	1.5715	1.0009	0.8756
21	0.9885	0.9931	1.5144	0.9984	0.8691
22	0.9911	0.9940	1.4265	0.9970	0.8636
23	0.9927	0.9945	1.3361	0.9958	0.8584
24	0.9930	0.9941	1.2952	0.9953	0.8528*
25	0.9922	0.9934	1.3342	0.9942	0.8430*
26	0.9943	0.9951	1.4204	0.9958	0.8347**
27	1.0072	1.0070	1.5049	1.0079	0.8432
28	1.0248	1.0239	1.5537	1.0245	0.8580
29	1.0432	1.0420	1.5596	1.0422	0.8821
30	1.0565	1.0552	1.5392	1.0553	0.9114

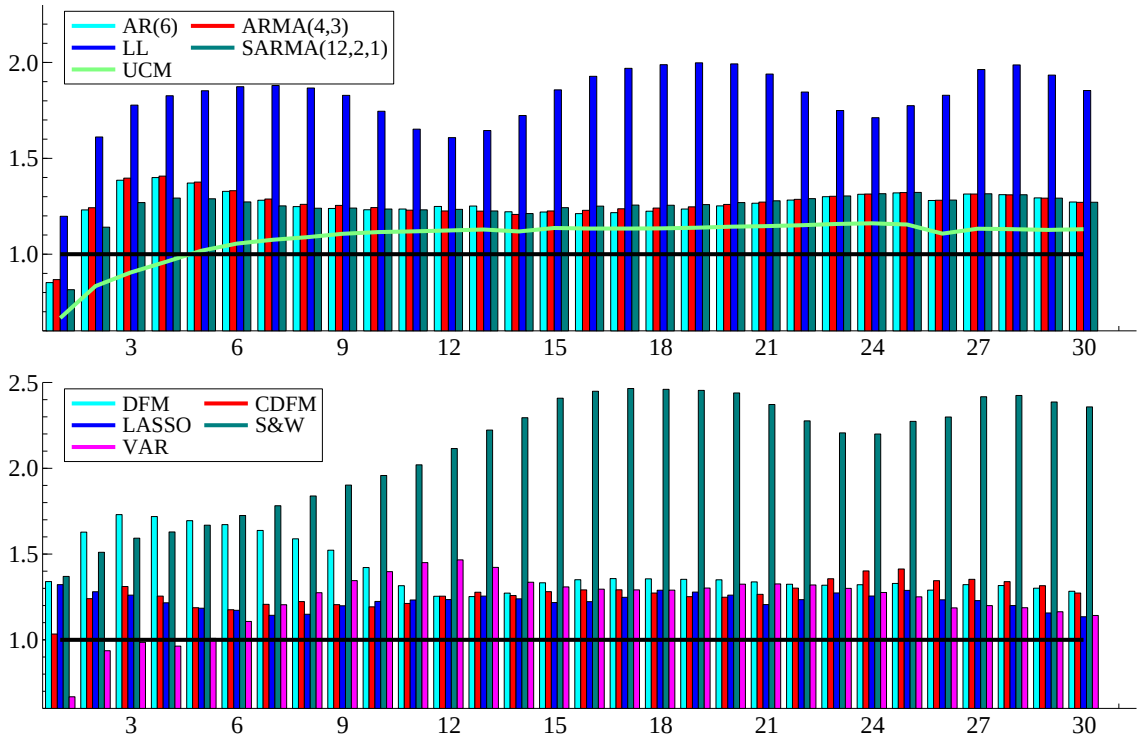
a VAR(2) model is more parsimonious than other models, thus it also has a lower estimation error which may benefit forecasting. However, the precision of the VAR forecasts deteriorates for longer horizons. It is highly encouraging that DFS produces a highly precise forecasts for long horizons. This is of crucial importance in the context of El Niño forecasting as it can facilitate a timely warning system. For $h = 1$, the null of equal predictive accuracy in terms of squared forecast errors between CDFM and DFS cannot be rejected, suggesting again the role of parsimony in forecasting, as both methods involve principal components and the modelling of dynamic factors. The automatic variable selection from LASSO is also a way of achieving parsimony, but its forecasting performance is close to CDFM except for some minor improvement from 20-step ahead on. With the chosen tuning parameter in LASSO regression, we find that the number of nonzero coefficients is similar to the number of factors we extract in CDFM, where both models start with 888 explanatory variables. S&W appears to be the worst-performing model for point forecast with RMSE_h more than double of DFS for almost all h. From the bottom panel of Figure 6 which gives the ratio of RMSE between different models and the DFS method, we can also see that the RMSE of S&W seems to increase against that of DFS, suggesting that it becomes less reliable as h increases. If we consider both CDFM and LASSO as some shrinkage forecasting model, the bottom panel of Figure 6 shows that the effect of shrinkage erodes as their performances converge to that of VAR or DFM.

It is easily observed that in general univariate models deliver better point forecasts than multivariate models from Table 2. For simple models such as AR(6) and ARMA$(4, 3)$, the DFS method is eclipsed for $h = 1$. AR(6) has $\text{RMSE}_1 = 0.3089$ that is 15% smaller than DFS, which is slightly bigger than SARMA$(12, 2, 1)$. Although they are simple in form, their ARMA lags are flexible enough and expected to accurately pick up concurrent behaviour of the Niño3.4 time series. From the top panel of Figure 6, SARMA$(12, 2, 1)$ seems to perform better than AR(6) and ARMA$(4, 3)$ only for short horizons, and they converge as h increases because for those models their point forecasts simply converge to the unconditional mean levels in the long term which are theoretically equal if ignoring estimation errors. The main reason for basing our DFS method on the UCM is that it delivers the most accurate point forecasts for all h among all models considered. Unlike ARMA models which converge to the unconditional mean, UCM decomposes Niño3.4 series and captures latent dynamics, i.e.

seasonal and cyclic processes, to produce more precise point forecast. For $h = 1$, 2 and 3, DFM significant outperforms the DFS method by 35%, 19% and 7% respectively. For $h = 3$ and 5, we fail to reject the null of equal predictive accuracy. This tells that our procedure may be not fully beneficial in the short term if the first stage univariate model is already well capable of producing accurate point forecasts. From $h = 6$ on however, significant improvements are made by coupling UCM with our DFS procedure. This improvement is also quite stable for all $h > 6$ as illustrated by the distance of two lines in the top panel of model 6. In contrast, the worst-performing LL which fully foregoes the stationary dynamics of hidden processes for Niño3.4 clearly implies some "term structure" of its forecasting ability, at an annual frequency in particular.

Figure 6

Ratio of RMSE between different models and UC_DFS



The panels show the ratio of RMSE between the selected models and the UC_DFS model for different lag length h. Top: univariate models. Bottom: multivariate models. A ratio higher than 1 means a better performance (lower RMSE) for the UC_DFS model.

Table 3 and 4 are similar to Table 1 and 2 but based on linex loss function. From the comparisons among multivariate models, it follows immediatelt that the VAR model performs well in predicting the direction of change as it dominates the DFS method for

more h-step forecasts than the case of squared error loss function in the short term. For $h \leq 6$, VAR significantly outperforms our DFS method except for $h = 6$ where the null of equal predictability is rejected. For one quarter or two ahead, forecasts for direction of change are advised to be produced by parsimonious models like VAR which leads to fewer underestimations. However, for long lead prediction, the linex loss of VAR increases similar to the case of RMSE. From the bottom panel of Figure 7 we observe yearly "term structure" of MLFE for all multivariate models except for LASSO, similar to the RMSE for LL and this phenomenon is the most visible for VAR which performs very poorly for h around 12. For the middle run, the HLM_DM test fails to tell the statistical difference between LASSO and the DFS method, which suggests that shrinkage employed by LASSO has similar effect in determining direction of change to the second stage DFM with $(v_t, X_t')'$ used in our DFS procedure. For univariate models, MLFE leads to conclusions similar to the case of RMSE. But the short-horizon performance for other univariate models are more observable than the case RMSE. Even the naive LL model is superior to the DFS method at $h = 1$ and tells no significant difference for $h = 2$, 3 and 5. The three ARMA models perform well for all h, especially for the SARMA(12,2,1) which dominates DFS for $h = 1$ to 4 and it is even slightly better than our first stage model UCM for $h > 15$ and the HLN_DM test cannot reject the null of equal predictability between it and the DFS for many h-step ahead forecasts. This interesting finding potentially implies that accounting for the seasonal variation of Niño3.4 plays an important role in forecasting the direction of change.

4.4 Forecasting results for the El Niño events

Table 5 summarizes the MRPS for some chosen models. We leave out the discussion of CDFM, LASSO, S&W because it is relatively more time-consuming to simulate from those models due to the fact that for different h one has a different model, which actually refrains us using these models to produce probability forecast. Contrary to the case of RMSE and MLFE, in the short horizons multivariate models seem to perform better in terms of ranked probability score loss. For $h \leq 3$, the winner is VAR. It gives MRPS of 0.09, 0.11 and 0.12 while those given by ARMA(4,3) which performs very well in terms of linex loss are 0.17, 0.19 and 0.20 respectively, almost doubling MRPS achieved by other models including our

Table 3

MLFE of Multivariate Models

The table reports the root mean linex forecast error $MLFE_h$ for multivariate benchmark models as explained in Section 4.1. A cell without a symbol (except for the column with header UC_DFS) means the DFS method outperforms another model at the 5% level. A shaded cell indicates that the UC_DFS underperforms compared to another model at the 5% level. Single asterisk "$*$" indicates that the UC_DFS model outperforms another model only at the 10% level. Double asterisks "$**$" indicates no significant difference in predictive accuracy between US_DFS and another model.

Step h	DFM	CDFM	LASSO	S&W	VAR	UC_DFS
1	0.0819**	0.0709**	0.1917	0.0713**	0.0274	0.0712
2	0.3428	0.2160	0.3085	0.1970	0.0940	0.1353
3	0.3929**	0.3976	0.4154	0.4223	0.1521	0.2170
4	0.5235*	0.5232	0.4845	0.7271	0.1959	0.3072
5	0.8587*	0.5985	0.5409	1.0239	0.2705**	0.3827
6	0.9654	0.6814	0.5739*	1.2486	0.3968	0.4365
7	0.8124	0.7748*	0.5596**	1.3748	0.5609*	0.4713
8	0.7627	0.8115*	0.5468**	1.4065	0.7472	0.4937
9	0.7591	0.7852	0.5523**	1.4142	0.9992	0.5085
10	0.7926	0.7880	0.5497**	1.4534	1.2208	0.5095
11	0.8496	0.8314	0.5339*	1.5317	1.4865	0.5127
12	0.8853	0.8900	0.5089**	1.7008	1.7418	0.5158
13	0.9407	0.9076	0.5235**	2.0355	1.5009	0.5141
14	1.0303	0.8747	0.5109*	2.5172	1.1947	0.5285
15	0.9840	0.8416*	0.5079**	2.9457	1.0798	0.5103
16	0.7557	0.8367**	0.5652**	3.0940	1.0583	0.5150
17	0.6029	0.8416*	0.6175**	2.9823	1.0488	0.5203
18	0.6240	0.8383**	0.7126	2.8015	1.0305	0.5210
19	0.7009	0.8270**	0.7004	2.7408	1.0304	0.5153
20	0.7568	0.8260*	0.6662	2.7628	0.9976	0.5121
21	0.7980	0.8362	0.6270	2.7266	0.9448	0.5105
22	0.8789*	0.8624	0.6154**	2.6746	0.9218	0.4940
23	0.9317	0.9015	0.6395	2.6479	0.8942	0.4784
24	1.0255	0.9333	0.6297	2.5907	0.8505	0.4634
25	1.1381	0.9428	0.6287**	2.4299	0.8016	0.4501
26	1.2564	0.9363	0.5881*	2.2459	0.7629	0.4904
27	1.3376	0.9427	0.5976	2.0900	0.7331	0.4459
28	1.1483	0.9526	0.6164	1.9108	0.7077	0.4494
29	0.8956	0.9385	0.6390	1.7181	0.6816	0.4629
30	0.7196	0.8846*	0.6794	1.5882	0.6683	0.4790

Table 4

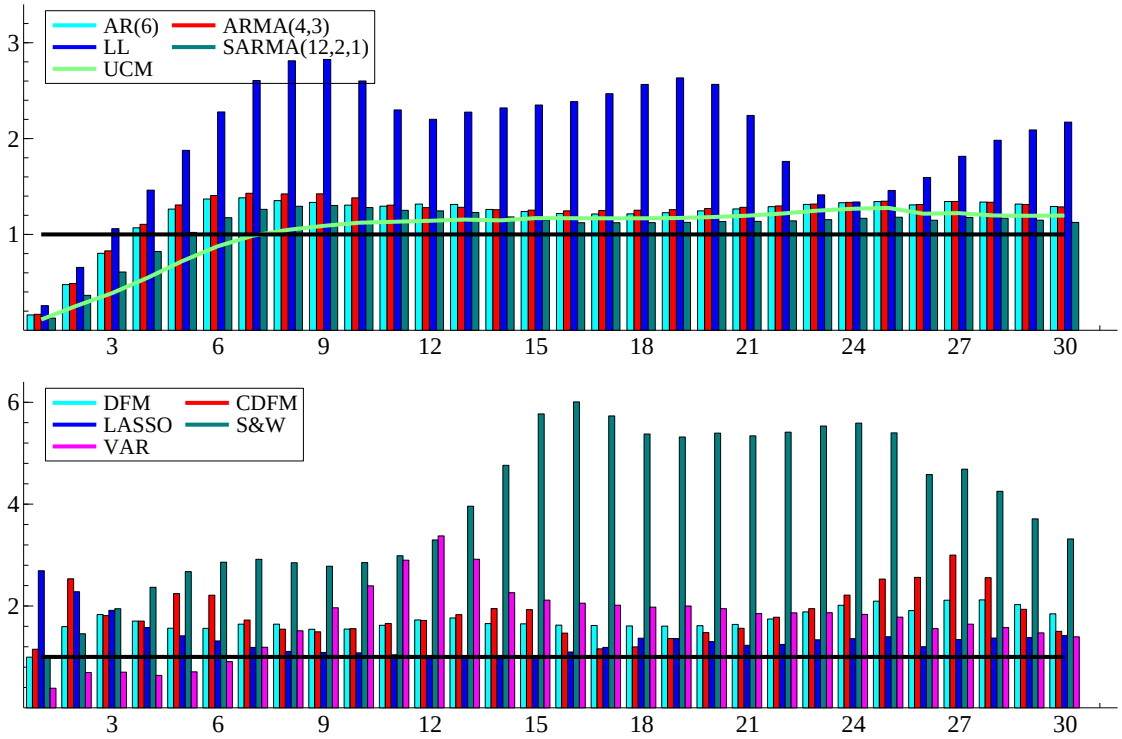
MLFE of Univariate Models

The table reports the root mean linex forecast error $MLFE_h$ for univariate benchmark models as explained in Section 4.1. A cell without a symbol (except for the column with header UC_DFS) means the DFS method outperforms another model at the 5% level. A shaded cell indicates that the UC_DFS underperforms compared to another model at the 5% level. Single asterisk "$*$" indicates that the UC_DFS model outperforms another model only at the 10% level. Double asterisks "$**$" indicates no significant difference in predictive accuracy between US_DFS and another model.

Step h	AR(6)	ARMA(4,3)	LL	SARMA(12,2,1)	UCM
1	0.0580	0.0606	0.0932	0.0441	0.0383
2	0.2330	0.2383	0.3204**	0.1699	0.1174
3	0.4799	0.4952**	0.6336**	0.3462	0.2176
4	0.7356**	0.7606**	1.0058	0.5388	0.3533**
5	0.9492**	0.9811**	1.4089**	0.7300**	0.5101
6	1.0786	1.1069	1.7934	0.8808	0.6497**
7	1.1224	1.1604	2.1160	0.9756	0.7540*
8	1.1200	1.1774	2.3266	1.0202	0.8165*
9	1.1074	1.1816	2.3448	1.0295	0.8505**
10	1.0804	1.1438	2.1526	1.0091*	0.8741*
11	1.0649	1.0743	1.8914	0.9806	0.8762
12	1.0791	1.0485	1.8039	0.9726	0.8809
13	1.0758	1.0508	1.8636	0.9585*	0.8898*
14	1.0434	1.0405	1.9188	0.9319*	0.8921
15	1.0085	1.0189	1.9107	0.8892**	0.8951
16	0.9918	1.0147	1.9425	0.8702**	0.8957
17	0.9850	1.0155	2.0046	0.8679	0.8932*
18	0.9812	1.0117	2.0709	0.8637**	0.8878
19	0.9803	1.0059	2.1049	0.8578**	0.8811
20	0.9824	1.0017*	2.0235	0.8516	0.8765*
21	0.9884	1.0019	1.7499	0.8452*	0.8781
22	0.9968	1.0027	1.3619	0.8401**	0.8857
23	1.0031	1.0059	1.0789	0.8392**	0.8973
24	1.0068	1.0092*	1.0125	0.8419**	0.8923*
25	1.0095	1.0129	1.0968	0.8448**	0.8934
26	1.0157	1.0190	1.2375	0.8494	0.8788
27	1.0301	1.0305	1.3912	0.8596*	0.8715
28	1.0460	1.0447	1.5499	0.8707*	0.8713
29	1.0614	1.0585	1.6852	0.8822	0.8963
30	1.0726	1.0687	1.8031	0.8905**	0.9253

Figure 7

Ratio of MLFE between different models and UC_DFS



The panels show the ratio of MLFE between the selected models and the UC_DFS model for different lag length h. Top: univariate models. Bottom: multivariate models. A ratio higher than 1 means a better performance (lower MLFE) for the UC_DFS model.

DFS method. The performance of predicting the probability for an El Niño event (and for the other two events) using SARMA(12,2,1) quickly wanes towards bigger h and SARMA becomes the worst-performing model for $h > 12$ which is in a stark contrast to the predictive accuracy in terms of linex loss. The DFS method does perform best for $h > 6$ and the second stage simulation and third stage ensemble series improve the first stage UC model from $h = 6$ onwards; we then observe a clear jump in the relative predictability in terms of ranked probability for UCM.

To complete the discussion of the DFS method, we also apply the three-step procedure but with different specifications to the El Niño dataset which serves as a robustness check for our proposed three-step forecasting procedure. The forecasting performance is evaluated via RMSE and results are reported in Appendix D.

Table 5

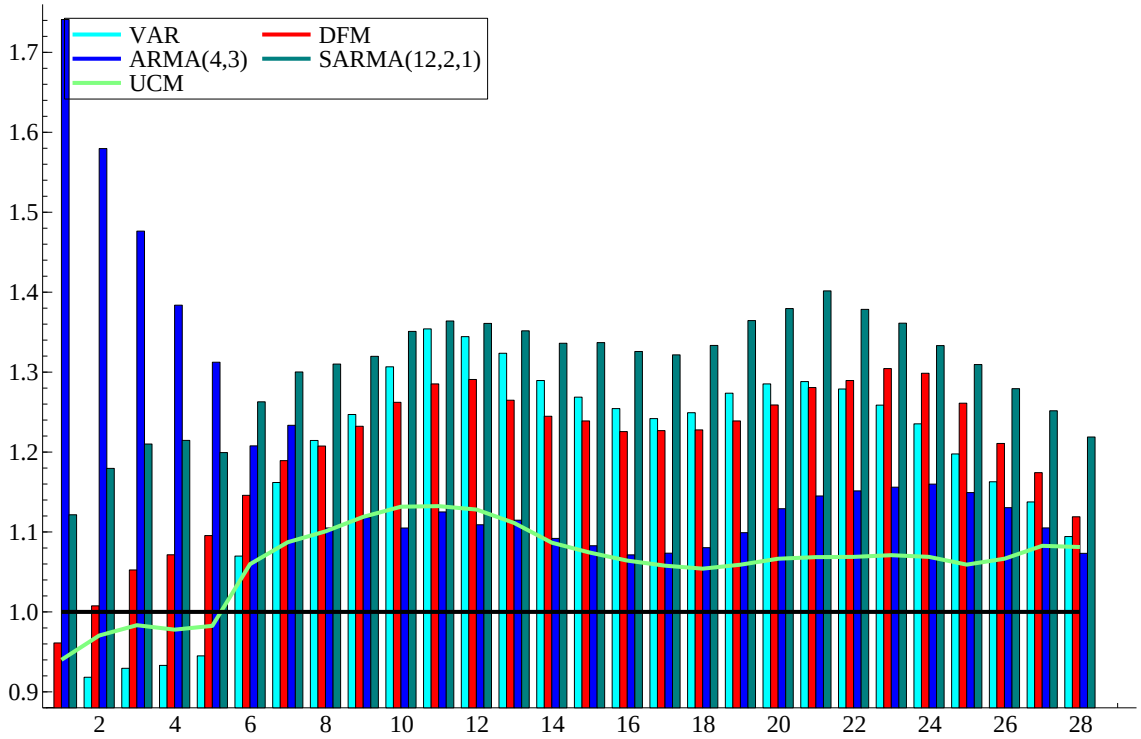
MRPS for a selection of Models

The table reports the root mean ranked probability score loss $MRPS_h$ for a selection of benchmark models as explained in Section 4.1. A cell without a symbol (except for the column with header UC_DFS) means the DFS method outperforms another model at the 5% level. A shaded cell indicates that the UC_DFS underperforms compared to another model at the 5% level. Single asterisk "$*$" indicates that the UC_DFS model outperforms another model only at the 10% level. Double asterisks "$**$" indicates no significant difference in predictive accuracy between US_DFS and another model.

Step h	VAR	DFM	ARMA(4,3)	SARMA(12,2,1)	UCM	UC_DFS
1	0.0884	0.0965	0.1748	0.1126	0.0977	0.1004
2	0.1079	0.1184**	0.1856	0.1386	0.1180**	0.1175
3	0.1239	0.1403*	0.1968	0.1613	0.1357**	0.1333
4	0.1396**	0.1603	0.2070	0.1817	0.1514*	0.1496
5	0.1564**	0.1813	0.2172	0.1985	0.1683**	0.1655
6	0.1746	0.1870	0.1971	0.2061	0.1790*	0.1632
7	0.1866	0.1910	0.1981	0.2088	0.1807	0.1606
8	0.1931	0.1920	0.1757	0.2083	0.1812	0.1590
9	0.1954	0.1931	0.1752	0.2068	0.1815	0.1567
10	0.2003	0.1935	0.1694	0.2071	0.1795	0.1533
11	0.2046	0.1942	0.1700	0.2061	0.1771	0.1511
12	0.2034	0.1953	0.1678	0.2059	0.1767	0.1513
13	0.2029	0.1939	0.1709	0.2072	0.1763	0.1533
14	0.2018	0.1948	0.1709	0.2091	0.1760	0.1565
15	0.2007	0.1960	0.1713	0.2115	0.1759	0.1582
16	0.2002	0.1956	0.1710**	0.2116	0.1758	0.1596
17	0.1977	0.1953	0.1709**	0.2104	0.1743*	0.1592
18	0.1975	0.1941	0.1708*	0.2108	0.1725*	0.1581
19	0.1978	0.1924	0.1707	0.2119	0.1702	0.1553
20	0.1951	0.1911	0.1714	0.2094	0.1676	0.1518
21	0.1918	0.1907	0.1705	0.2087	0.1647	0.1489
22	0.1926	0.1942	0.1734	0.2076	0.1666	0.1506
23	0.1927	0.1997	0.1770	0.2084	0.1697	0.1531
24	0.1932	0.2031	0.1814	0.2085	0.1730	0.1564
25	0.1757	0.1850	0.1686	0.1921	0.1608	0.1467
26	0.1578	0.1643	0.1534	0.1736	0.1498	0.1357
27	0.1397	0.1442	0.1357	0.1537	0.1376	0.1228
28	0.1195*	0.1222	0.1172*	0.1331	0.1222	0.1092

Figure 8

Ratio of MRPS between different models and UC_DFS



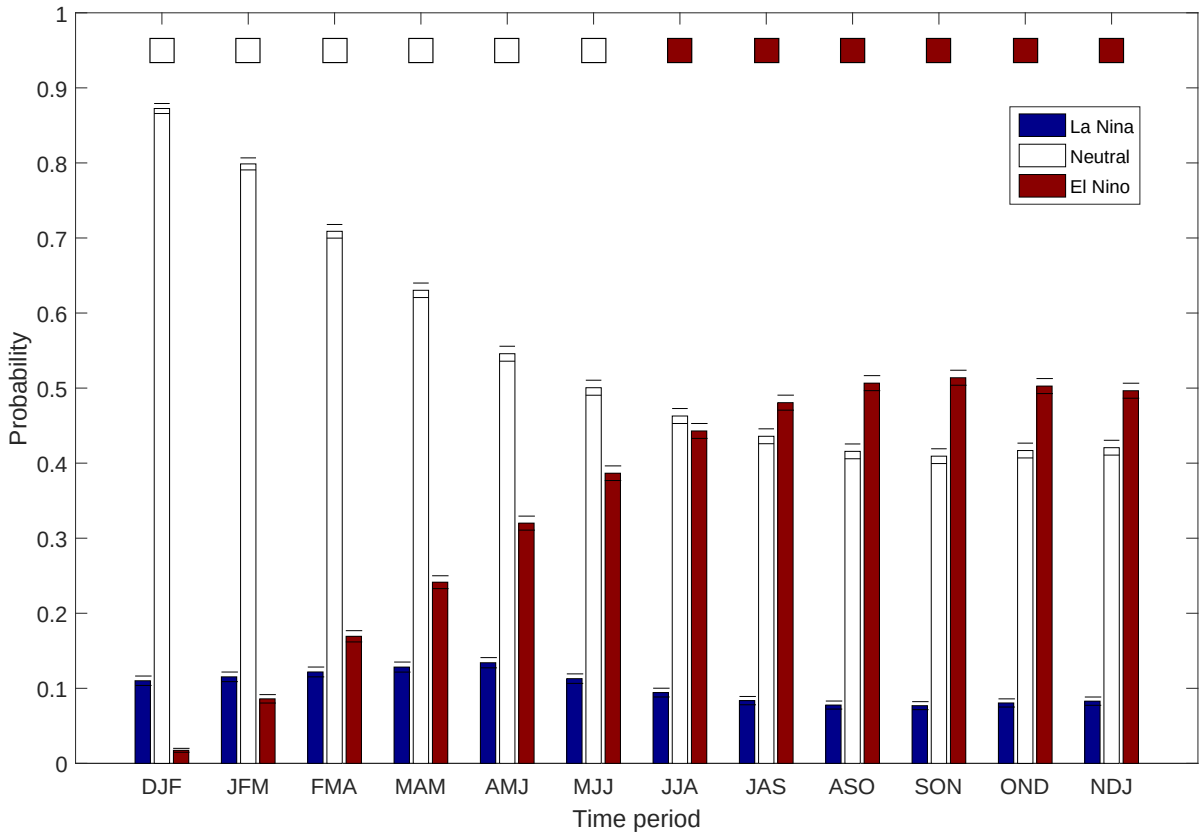
The figure shows the ratio of mean rank probability scores (MRPS) between the selected models and the UC_DFS model for different lag length h. A ratio higher than 1 means a better performance (lower MRPS) for the UC_DFS model. The probability forecasts that are the input for the rank probability scores come from simulated forecasts. All models are cast into state space form from which multi-step ahead forecasts are drawn using a simulation smoother, treating future values as missing.

5 Conclusions

We propose a novel procedure for the long term forecasting of the Niño3.4 time series and the related warm El Niño events. The time series has a complex dynamic structure and relies on a potentially large set of explanatory variables with many lagged dependencies as well. In our proposed forecasting procedure, a selection procedure to determine which explanatory variables and for which lags need to be included in the model is not needed. The dynamic factor model is adopted to produce simulated paths of prediction errors that are conditional on all explanatory variables. Hence all explanatory variables make a contribution but the simulations will consider a variation of different paths that may affect the forecasts differently. It turns out that the average of these forecasts are highly precise, even for longer lead forecasts. A wide range of empirical evidence with various robustness verifications is

Figure 9

Monthly Three-Months Period Probability Forecasts for 2009, dated 11-2008 (1-15 steps ahead forecasts)

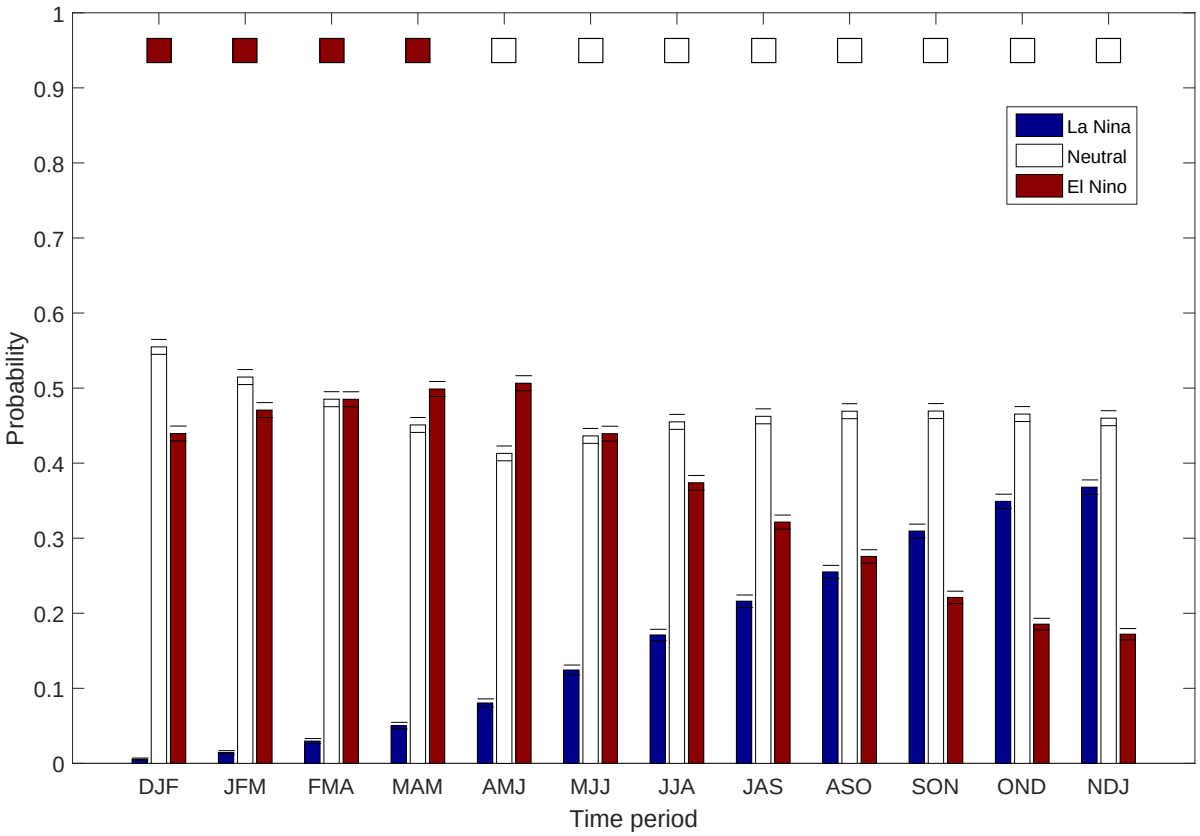


The figure shows monthly three-months period probability forecasts for the whole year 2009. The squares in the top of the figure show the observed categories. The first half year shows no El Nino or La Nina event with the model assigning the most probability to no event. The second half of the year we have 6 El Nino events (possibly more in 2010) and the model strongly increases the probability of an El Nino event.

provided. Amongst other avenues for research, we may explore the density forecasts that can be made part of our simulation-based methodology. Furthermore, the implications of our methodology for the forecasting of El Niño events can be investigated in settings that are familiar to climate research.

Figure 10

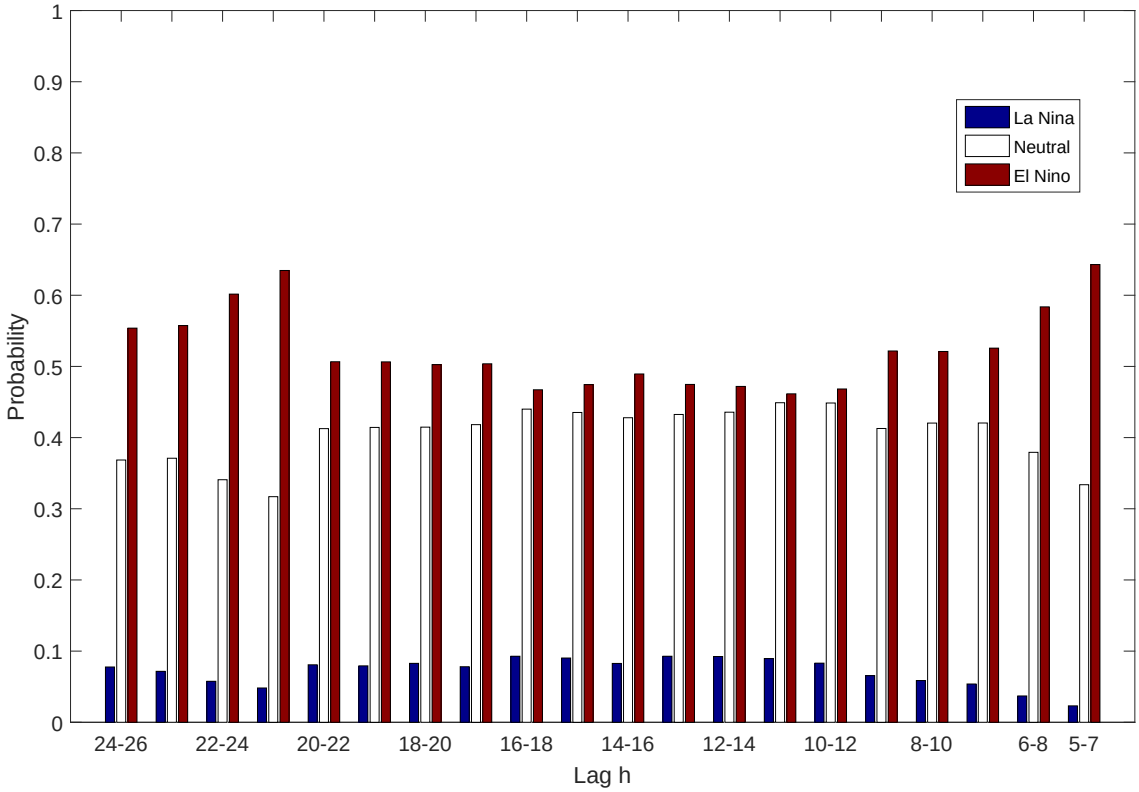
Monthly Three-Months Period Probability Forecasts for 2005, dated 11-2004 (1-15 steps ahead forecasts)



The figure shows monthly three-months period probability forecasts for the whole year 2005. The squares in the top of the figure show the observed categories. The first four periods show an El Nino event with the model assigning most of the probability to an El Nino event and no event. The remaining 8 periods of the year have no events which is the largest forecasted category by the model.

Figure 11

Probabilistic forecast for the El Nino event of DJF 2010 for 5-26 steps ahead

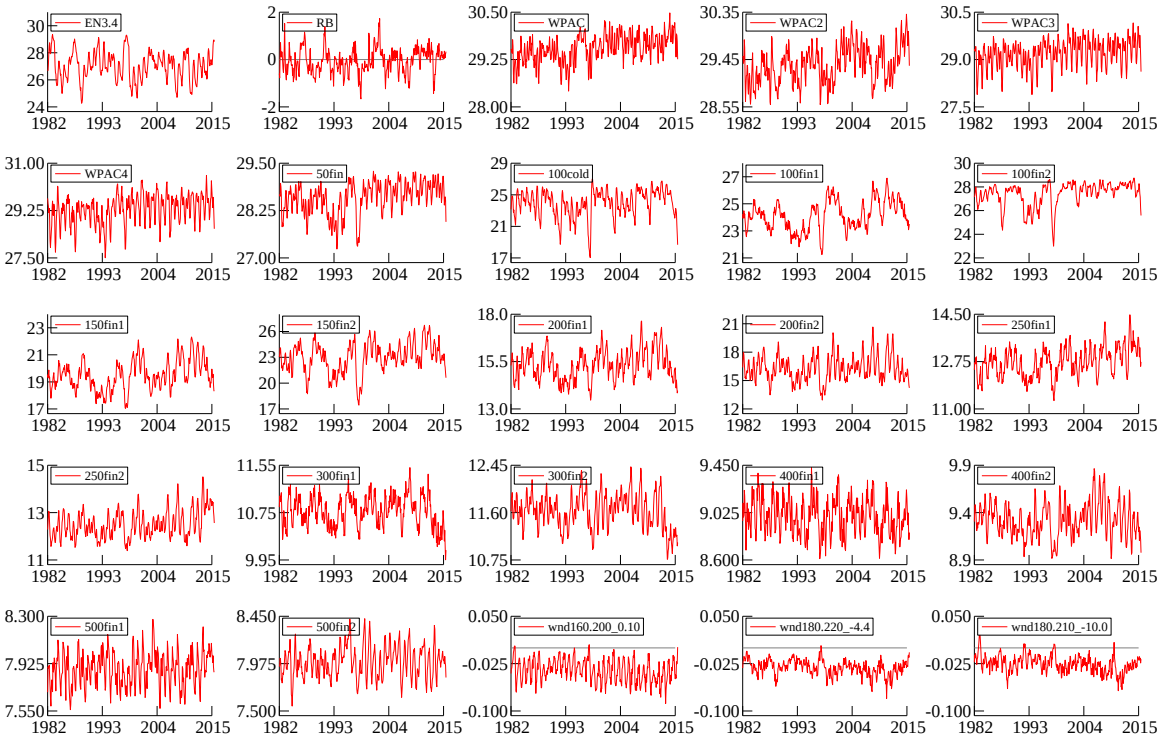


The figure shows the simulated probability forecast for each of the categories and for different lag length h leading up to the El Nino event of DJF (2010).

Appendices

A Data

Figure 12
Data Graphs



The figure presents the time series plots of all variables in our Data Set.

Table 6

Details of Data Set

Description of the dependent and explanatory regression variables used in our study. The acronyms, the variable types, the regions of where the variables are measured and calculated, and the sources from which the variables are obtained. The sources are the NOAA-OI-SST-V2 database (OISST) as provided by NOAA/OAR/ESRL PSD (see `http://www.esrl.noaa.gov/psd/`), the NCEP/NCAR reanalysis by Kalnay, Kanamitsu, Kistler, Collins, Deaven, Gandin, Iredell, Saha, White, Woollen, Zhu, Leetmaa, Reynolds, Chelliah, Ebisuzaki, Higgins, Janowiak, Mo, Ropelewski, Wang, Jenne, and Joseph (1996), NCEP, Subsurface Temperature And Salinity Analyses (ISHII) by Ishii et al. (2005), as archived at the National Center for Atmospheric Research, Computational and Information Systems Laboratory (see `http://www.rda.ucar.edu/datasets/ds285.3/`), and, finally, the Hadley Centre EN4.0.2 (EN4) as analysed by Good, Martin, and Rayner (2013).

	Acronym	**Variable type**	**Region**	**Source**
	Niño3.4	sea surface temp.	$[190e - 240e] \times [5s - 5n]$	OISST
1	RB	sea surface temp.	$[180e - 280e] \times [65s - 50S]$	OISST
2	WPAC	sea surface temp.	$[140e - 160e] \times [5s - 5n]$	OISST
3	WPAC2		$[140e - 180e] \times [10s - 5n]$	OISST
4	WPAC3		$[120e - 170e] \times [10s - 5n]$	OISST
5	WPAC4		$[140e - 160e] \times [10s - 0]$	OISST
6	Temp.50m	subsurface temp.	$[120e - 170e] \times [10s - 5n]$	ISHII, EN4
7	Temp.100m "cold"		$[140e - 210e] \times [5n - 10n]$	ISHII, EN4
8	Temp.100m R1		$[120e - 140e] \times [10s - 5n]$	ISHII, EN4
9	Temp.100m R2		$[150e - 180e] \times [7s - 7n]$	ISHII, EN4
10	Temp.150m R1		$[120e - 140e] \times [10s - 5n]$	ISHII, EN4
11	Temp.150m R2		$[150e - 180e] \times [7s - 7n]$	ISHII, EN4
12	Temp.200m R1		$[120e - 140e] \times [10s - 7n]$	ISHII, EN4
13	Temp.200m R2		$[150e - 180e] \times [7s - 7n]$	ISHII, EN4
14	Temp.250m R1		$[120e - 140e] \times [7s - 7n]$	ISHII, EN4
15	Temp.250m R2		$[140e - 170e] \times [7s - 7n]$	ISHII, EN4
16	Temp.300m R1		$[120e - 140e] \times [7s - 7n]$	ISHII, EN4
17	Temp.300m R2		$[160e - 200e] \times [10s - 3n]$	ISHII, EN4
18	Temp.400m R1		$[120e - 140e] \times [5s - 5n]$	ISHII, EN4
19	Temp.400m R2		$[150e - 170e] \times [10s - 0]$	ISHII, EN4
20	Temp.500m R1		$[120e - 140e] \times [5s - 5n]$	ISHII, EN4
21	Temp.500m R2		$[150e - 170e] \times [10s - 0]$	ISHII, EN4
22	WND1	zonal wind stress	$[180e - 220e] \times [4s - 4n]$	NCEP
23	WND2		$[180e - 210e] \times [10s - 0]$	NCEP
24	WND3		$[160e - 200e] \times [0 - 10n]$	NCEP

B Two key models in our forecasting procedure

B.1 Univariate unobserved components time series models

Consider an observed monthly time series y_t that we assume is generated by the dynamic linear model consisting of a set of unobserved components as given by (1), that is

$$y_t = \mu_t + \gamma_t + \sum_{j=1}^{3} \psi_{j,t} + \varepsilon_t, \qquad t = 1, \ldots, T,$$

where μ_t represents the trend component, γ_t represents the monthly seasonal component, $\psi_{j,t}$ represents a stochastic cyclical process with certain amplitude and frequency, for $j = 1, 2, 3$, and ε_t represents the disturbance or noise component. In a linear Gaussian model, we assume that these six unobserved components are generated by linear dynamic stochastic processes depending on Gaussian noise terms and are generated independently of each other. In case of the trend component, we assume a simple random walk process, that is

$$\mu_{t+1} = \mu_t + \eta_t, \qquad \eta_t \sim \mathbb{N}(0, \sigma_\eta^2), \qquad t = 1, \ldots, T, \tag{10}$$

where the initial value μ_1 is treated as an diffuse prior variable, that is $\mu_1 \sim N(0, k)$ where $k \to \infty$. The variance σ_η^2 is unknown and is treated as a parameter that will be estimated by the method of maximum likelihood. A comprehensive discussion on these issues is provided in Durbin and Koopman (2012, Chapter 2). The seasonal component γ_t can be viewed as a set of 11 dummies that capture the average monthly deviations from the trend component. The twelfth dummy is restricted to ensure that the sum of the 12 dummies equals zero. These monthly dummies can stochastically vary over time through similar random walk specifications as in (10). The overall seasonal time-variation is determined by the variance σ_γ^2. The cycle components are generated by stationary autoregressive moving average (ARMA) processes with autoregressive polynomials that have complex roots. In the formulation of Harvey (1989), and others, the complex roots are imposed on the autoregressive polynomial.

We then formulate the j-th, for $j = 1, 2, 3$, cyclical stochastic process as

$$\begin{pmatrix} \psi_{j,t+1} \\ \psi^*_{j,t+1} \end{pmatrix} = \phi_j \begin{bmatrix} \cos(\lambda) & \sin(\lambda) \\ -\sin(\lambda) & \cos(\lambda) \end{bmatrix} \begin{pmatrix} \psi_{j,t} \\ \psi^*_{j,t} \end{pmatrix} + \begin{pmatrix} \kappa_{j,t} \\ \kappa^*_{j,t} \end{pmatrix}, \quad \kappa_{j,t}, \kappa^*_{j,t} \sim \mathbb{N}(0, \sigma^2_{\kappa,j}), \tag{11}$$

for $t = 1, \ldots, T$, where the initial values for both elements of $\psi_{j,1}$ and $\psi^*_{j,1}$ are treated as independent Gaussian random variables with their moments set equal to the unconditional mean and variance of $\psi_{j,t}$ and $\psi^*_{j,t}$, respectively, that is $\psi_{j,1}, \psi_{j,1} \sim \mathbb{N}[0, \sigma^2_\kappa / (1 - \phi^2_j)]$, for $j = 1, 2, 3$. The coefficients for persistence, ϕ_j, frequency, λ_j and variance, $\sigma^2_{\kappa,j}$ (that determines the amplitude of cycle $\psi_{j,t}$), are treated as unknown parameters, for $j = 1, 2, 3$. It can be shown that the cyclical process $\psi_{j,t}$ as formulated in (11) can be represented as an autoregressive moving average process of order (2,1), that is ARMA(2,1), where the second-order autoregressive polynomial can only have stationary, complex roots under the parameter restrictions $0 < \phi_j < 1$, $0 < \lambda_j < 2\pi$ and $\sigma^2_{\kappa,j} > 0$. Finally, we assume that the disturbance component ε_t is assumed to be a Gaussian white noise process, we simply have $\varepsilon_t \sim \mathbb{N}(0, \sigma^2_\varepsilon)$, for $t = 1, \ldots, T$. All Gaussian disturbances are mutually and serially uncorrelated from each other. The model properties are further discussed in Durbin and Koopman (2012, Ch. 3).

B.2 Dynamic Factor Model

To analyse, model and forecast the observed time series for some key variable y_t, it may be beneficial to incorporate data information from a time series panel of related variables which we denote by X_t, for $t = 1, \ldots, T$. In our study, we will treat y_t as a scalar variable and X_t as $(N - 1) \times 1$ vector. Although y_t may be subject to non-stationary dynamics, we assume that all individual time series in X_t are weakly stationary. The basic dynamic factor model for the $N \times 1$ joint vector $(y_t, X'_t)'$ is given

$$\begin{pmatrix} y_t \\ X_t \end{pmatrix} = \begin{bmatrix} \lambda' \\ \Lambda_X \end{bmatrix} f_t + \xi_t, \qquad f_t = \Phi f_{t-1} + \nu_t, \tag{12}$$

for $t = 1, \ldots, T$, where λ is the $r \times 1$ vector of loadings, Λ_X is the $(N - 1) \times r$ factor loading matrix, $\Lambda = [\lambda \,,\, \Lambda'_X]'$ is the $N \times r$ factor loading matrix, with A' being the transpose of matrix A, f_t is the $r \times 1$ vector with latent dynamic factors, ξ_t is the $N \times 1$ observation disturbance

vector, which is assumed to be normally distributed, Φ is the $r \times r$ autoregressive coefficient matrix, and ν_t is the $r \times 1$ normally distributed factor disturbance vector. The dynamic factor f_t represents the common dynamic variations in all time series variables in y_t and X_t. The dynamic process for f_t is specified as a strictly stationary vector autoregressive process. Hence, matrix Φ is subject to the appropriate conditions for stationarity. Other stationary, linear dynamic processes can also be considered for f_t. For identification purposes we further assume that the factors are normalized, that is $\mathrm{E}(f_t) = 0$ and $\mathrm{Var}(f_t) = I_r$ with I_k being the $k \times k$ identity matrix for any positive integer k. The disturbance vectors ξ_t and ν_t are normally distributed with variance matrices Σ_ξ and Σ_ν, respectively, and are assumed to be mutually and serially uncorrelated for all time periods. To enforce the normalization of the factors f_t, the variance matrix Σ_ν is restricted to be $\Sigma_\nu = I_r - \Phi\Phi'$. In our treatment for this study, the initial factor f_1 is treated as a random vector with mean zero and unity variance matrix; this is conform the normalization assumption for all factors f_t, for $t = 1, \ldots, T$.

The unknown coefficient matrices Λ, Σ_ξ and Φ are functions of a parameter vector that we denote by ψ. This dynamic factor model is stylized for the purpose of presenting our developments below. However, our results and developments are general and can be applied to other dynamic factor model specifications. The dynamic factor model (12) provides a convenient framework for obtaining simple descriptions of potentially complex interactions between the variables. In particular, the common factors summarize partly the commonalities in the dynamic variations in the related variables X_t. Furthermore, the factors deliver a parsimonious description of the relationships between the variables of interest in y_t and the the related variables in X_t. The dynamic factor model is mainly used to approximate the true and unknown data generation process. It is not intended to be an exact representation of the true underlying data generation process.

The estimation of the unknown parameter ψ can take place by the method of maximum likelihood but the high dimension of ψ typically complicates and slows down the numerical optimisation process. For this reason, we adopt the two-step estimation procedure of Doz et al. (2011); this straightforward estimation method provides a consistent estimate of ψ in a computational fast manner and without using a numerical optimisation procedure. The first step is based on the r principal components, denoted by $\widetilde{f}_t$, and computed from the data matrix $[(y_1, X_1')', \ldots, (y_T, X_T')']'$, after standardisation. The r principal components

correspond to the r largest eigenvalues of the sample variance matrix of the data matrix. By replacing f_t with $\widetilde{f}_t$ in (12), we obtain the two linear regression equations

$$\begin{pmatrix} y_t \\ X_t \end{pmatrix} = \Lambda \widetilde{f}_t + \xi_t^*, \qquad \widetilde{f}_t = \Phi \widetilde{f}_{t-1} + \nu_t^*, \tag{13}$$

where we assume that the disturbances ξ_t^* and ν_t^* have the same properties as ξ_t and ν_t, respectively, of the original model (12). Hence we treat the r principal components as proxies for the r dynamic factors in f_t. By applying least squares to both linear regression equations, we obtain estimates for Λ, Σ_ξ and Φ which we denote by $\widehat{\Lambda}$, $\widehat{\Sigma}_\xi$ and $\widehat{\Phi}$, respectively. This completes the first step. The second step considers the model

$$\begin{pmatrix} y_t \\ X_t \end{pmatrix} = \widehat{\Lambda} f_t + \xi_t^\dagger, \qquad f_t = \widehat{\Phi} f_{t-1} + \nu_t^\dagger, \tag{14}$$

where we assume that the disturbances $\xi_t^\dagger$ and $\nu_t^\dagger$ are normally distributed with zero means and variance matrices $\widehat{\Sigma}_\xi$ and I_r, respectively. A dynamic factor analysis can be fully based on the model (14) with the use of the Kalman filter and related methods. This includes the forecasting of y_t multi-steps ahead but also the simulation of f_t conditional on all observations for $y_1, \ldots, y_T$ and $X_1, \ldots, X_T$, see Durbin and Koopman (2012, Chapter 4) for a detailed discussion.

C Technical justification of our forecasting procedure

Multivariate time series models can often be written in state space form and the dynamic factor model is a particular example of this. For detailed discussion on state space representation of time series models, readers can consult Harvey and Shephard (1993) and Durbin and Koopman (2012).

It is thus reasonable to assume the following DGP for a N-dimensional vector $(y_t, X_t')'$ where $y_t \in \mathbb{R}$ is the variable to forecast and $X_t \in \mathbb{R}^{N-1}$ collects other variables,

$$\begin{aligned} \begin{bmatrix} y_t \\ X_t \end{bmatrix} &= Z f_t + \epsilon_t, \quad \epsilon_t \sim N(0, H), \\ f_{t+1} &= T f_t + \eta_t, \quad \eta_t \sim N(0, Q), \end{aligned} \tag{15}$$

for $t = 1, \ldots, T$, where $f_t \in \mathbb{R}^p$ with $p < N$ is the dynamic factor with the transition matrix T subject to the usual stationarity restriction. The loading matrix $Z \in \mathbb{R}^{N \times p}$ and the covariance matrices $H \in \mathbb{R}^{N \times N}$ and $Q \in \mathbb{R}^{p \times p}$ are all constant matrices. T, Z, H and Q can contain unknown system parameters which normally requires estimation or calibration before one can forecast using this model. Forni, Hallin, Lippi, and Reichlin (2000) and Doz et al. (2011) term model (15) the exact dynamic factor model if the observation covariance matrix H is diagonal, and the approximate dynamic factor model if H is a band matrix with off-diagonal elements becoming smaller in absolute value if they are further away from the diagonal. If T is an identity matrix, i.e. f_t is a vector random walk, Harvey (1989) shows that in such a case there exists $N - p$ cointegration polynomials among $(y_t, X_t')'$. In the following, we assume an exact dynamic factor model with $H = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_N^2)$, and a stationary factor process, i.e. the biggest eigenvalue of T in absolute value is smaller than unity.

In many multivariate time series models such VAR, H is a zero matrix, but it is not hard to find cases where such a simplification is problematic. For example, people usually have little knowledge on which variables should be included in the VAR and which should not. Alternatively, in macroeconomics variables like GDP, inflation rate and price level are often observed with noise and measurement error. So with minor extension, a state space model such as model (15) serves as a closer description of the DGP. Since y_t is the only variable

that we are interested, one option that practitioners often choose is to fit a univariate model for y_t. One major reason for doing so is to keep model parsimonious such that negative effects resulted from issues such as misspecification and estimation errors are minimized. To demonstrate, let us partition $Z = (Z_1', Z_2')'$ with Z_1 being the first row of Z and Z_2 being Z without the first row, so from (15) we can equivalently write a univariate DGP for y_t as

$$\begin{aligned} y_t &= Z_1\kappa\kappa' f_t + Z_1(I_p - \kappa\kappa')f_t + \epsilon_{1,t}, \quad \epsilon_{1,t} \sim N(0, \sigma_1^2), \\ \kappa'\kappa\kappa' f_{t+1} &= \kappa' T\kappa\kappa' f_t + \kappa' T(I_p - \kappa\kappa')f_t + \kappa'\eta_t - \kappa'(I_p - \kappa\kappa')f_{t+1} \\ &= \kappa' T\kappa\kappa' f_t + \kappa' T(I_p - \kappa\kappa')f_t + \kappa'\eta_t. \end{aligned}$$

A univariate model for y_t implies that the weighting matrix κ belongs to the Stiefel manifold, i.e. $\kappa \in \{k \in \mathbb{R}^{p\times q} : k'k = I_q\}$ with $q \leq p$. By specifying a certain structure of $l = Z_1\kappa$ and $T^* = \kappa' T\kappa$, we have a univariate model which can be identified and estimated. It follows that

$$\begin{aligned} y_t &= l g_t + \epsilon_t^*, \\ g_{t+1} &= T^* g_t + \eta_t^*, \end{aligned} \tag{16}$$

where the q-dimensional signal or state is a linear combination of the p-dimensional factor, or $g_t = \kappa' f_t$. Moreover, it is necessary to consider κ as a random matrix due to the choice of the univariate model, the dimension of the state g_t resulted from this chosen model and the stochastic nature of the original factor process f_t. Therefore, the requirement $E(\kappa\kappa') = I_p$ is needed and κ belongs to a restricted Stiefel manifold, i.e.

$$\kappa \in \{k \in \mathbb{R}^{p\times q} : k'k = I_q, E(kk') = I_p\}. \tag{17}$$

Consequently, the observation and state errors are

$$\begin{aligned} \epsilon_t^* &= Z_1(I_p - \kappa\kappa')f_t + \epsilon_{1,t}, \qquad \epsilon_t^*|\kappa \sim N(0, Z_1(I_p - \kappa\kappa')\Sigma_{f,tt}(I_p - \kappa\kappa')'Z_1' + \sigma_1^2), \\ \eta_t^* &= \kappa' T(I_p - \kappa\kappa')f_t + \kappa'\eta_t, \quad \eta_t^*|\kappa \sim N(0, \kappa'(T(I_p - \kappa\kappa')\Sigma_{f,tt}(I_p - \kappa\kappa')'T' + Q)\kappa), \end{aligned} \tag{18}$$

where $\Sigma_{f,tt}$ is the t-th diagonal block submatrix of the covariance matrix of $vec(F_T)$ with $F_t = (f_1', f_2', ..., f_t')'$. This shows that an exact DFM (15) implies a univariate model (16)

for y_t. Equivalently, a univariate model can be some reduced form of an exact DFM. For example, Durbin and Koopman (2012) show that an ARIMA model for y_t has $l = (1, 0, ..., 0)$, and in an unobserved component model, l contains zeros and ones. One should notice that from $l = Z_1\kappa$, Z_1 in (16) usually cannot be identified because κ is unknown and not unique. And also, the measurement equation resembles a lot the endogeneity problem as $g_t \not\perp \epsilon_t^*$.

To illustrate the idea, let us consider the following simple DGP for y_t with $x_t \in \mathbb{R}$ and a univariate factor process f_t

$$\begin{aligned} \begin{bmatrix} y_t \\ x_t \end{bmatrix} &= \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix} f_t + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}, \\ f_{t+1} &= \phi f_t + \eta_t, \end{aligned} \tag{19}$$

where $|\phi| < 1$ and $\epsilon_{1,t}$, $\epsilon_{2,t}$ and η_t are i.i.d normal random variable with variance $\sigma^2_{\epsilon_1}$, $\sigma^2_{\epsilon_2}$ and σ^2_η respectively. It is easy to show that the reduced form of y_t is an ARMA(1,1) model because

$$y_t = \phi y_{t-1} + \Lambda_1 \eta_{t-1} + \epsilon_{1,t} - \phi\epsilon_{1,t-1}. \tag{20}$$

So the univariate model for y_t can also be written as $y_t = \rho y_{t-1} + \theta\xi_{t-1} + \xi_t$ with $\xi_t \sim N(0, \sigma^2_\xi)$ where $\rho = \phi$, and θ and σ^2_ξ can be calculated by equalizing the first two autocovariance functions of y_t from model (19) and from the reduced model. Following the procedure of Durbin and Koopman (2012, Chapter 3), the univariate ARMA(1,1) model for y_t can be easily cast into a (non-unique) state space model just like (16).

The first stage of the proposed method implies a recursive filter of y_t. In case of Kalman filter, $E(g_t|y_{1:t-1}) = \sum_{j=1}^{t-1} w_{jt} y_j$ with filtering weights w_{jt}. Similar to many other linear univariate forecasting models which give $\hat{y}_t = E(y_t|y_{1:t-1}) = \sum_{j=1}^{t-1} w_{jt} y_j$, the weights w_{jt} for $j = 1, ..., t-1$ are only functions of $y_{1:t-1}$ with the residual or prediction error given by $(v_t = y_t - \hat{y}_t) \perp \hat{y}_t$. This orthogonality tells that based on a chosen univariate model, the prediction error v_t typically does not play any role in delivering point forecast and what matters is the estimated signal $\hat{g}_t = E(g_t|y_{1:t-1})$. It can already be seen from (18) that v_t may still contain information on dynamics of y_t, because the part related to $Z_1(I_p - \kappa\kappa')f_t$ equals zero only in expectation, and for a chosen univariate model it can be retrieved with the second stage of our method, i.e. applying a DFM for $(v_t, X_t')'$.

In many cases, the residual or prediction error $v_t = M_t(Y_{t-1})Y_t$, where $Y_j = (y_1, y_2, ..., y_j)'$ and $M_t(Y_{t-1})$ is the t-th row of a (near-)projection matrix $M(Y_{t-1})$ based on Y_{t-1} and weights mentioned in the previous paragraph. In ARMA models, regression analysis shows

$$M(Y_{t-1}) = I - \dot{Y}_{t-1}(\dot{Y}'_{t-1}\dot{Y}_{t-1})^{-1}\dot{Y}_{t-1},$$

where $\dot{Y}_{t-1}$ is some data matrix arranged from Y_{t-1} and different if the orders of ARMA specification are different. In linear state space models, we have

$$M(Y_{t-1}) = I - ZL(Y_{t-1})K(Y_{t-1}), \tag{21}$$

where $L(Y_{t-1})$ and $K(Y_{t-1})$ are only functions of Y_{t-1} if assuming there is no estimation error for Z and T. See Durbin and Koopman (2012, Chapter 4) for details of the stacked form Kalman filter. Our proposed method uses one-step prediction error $v_t = M_t(Y_{t-1})Y_t = y_t - l\hat{g}_t$ instead of the smoothing error $y_t - \sum_{j=1}^{T} W_{jt}y_j$ with the j-th smoothing weight W_{jt} being a function of all available data $y_{1:T}$ or Y_T. This is crucial because v_t is only a function of Y_{t-1} without y_t. It is then only a function of F_{t-1}, leaving information about f_t unexplained. It follows that

$$\begin{aligned} v_t &= Z_1\kappa\kappa' f_t + Z_1(I_p - \kappa\kappa')f_t + \epsilon_{1,t} - l\sum_{j=1}^{t-1} w_{jt}y_j \\ &= Z_1(I_p - \kappa\kappa')f_t + \epsilon_{1,t} + l(g_t - \sum_{j=1}^{t-1} w_{jt}y_j) \\ &= Z_1(I_p - \kappa\kappa')f_t + \hat{\epsilon}_{1,t}. \end{aligned}$$

The above makes clear two points. Firstly, the first part of the prediction error $Z_1(I_p - \kappa\kappa')f_t$ may lead to poor forecasting performance because part of the information related to the dynamics of f_t is not used when applying a chosen univariate model which is often misspecified. Additionally, the second part of the prediction error $\hat{\epsilon}_{1,t} = \epsilon_{1,t} + l(g_t - \sum_{j=1}^{t-1} w_{jt}y_j)$ results from estimating a reduced model (16) for y_t.

To choose a univariate model is to choose κ from a subspace of the Stiefel manifold $\{k \in \mathbb{R}^{p\times q} : k'k = I_q\}$, because a univariate model that specifies zero mean error terms or

$E(v_t) = 0$ actually requires $E(\kappa\kappa') = I_p$, making κ stochastic. For example, the Kalman filter gives zero mean prediction errors. Regression models give zero mean residuals and zero mean error term is always assumed. An ARMA model chosen based on Box-Jenkins procedure also leads to $E(v_t) = 0$. So together with the requirement $E(M(Y_{t-1})Y_t) = 0$ in general, one effectively needs $\kappa \in \{k \in \mathbb{R}^{p\times q} : k'k = I_q, E(kk') = I_p\}$. Yet κ is still undetermined as the realisation of κ also depends on realisation of F_T, i.e. the data Y_T.

In the second stage, the following DFM is formed

$$\begin{aligned} \begin{bmatrix} v_t \\ X_t \end{bmatrix} &= \begin{bmatrix} Z_1(I_p - \kappa\kappa') \\ Z_2 \end{bmatrix} f_t + \begin{bmatrix} \hat{\epsilon}_{1,t} \\ \epsilon_{2,t} \end{bmatrix}, \\ f_{t+1} &= T f_t + \eta_t, \quad \eta_t \sim N(0, Q). \end{aligned} \tag{22}$$

Go back to the illustration of model (19). If we choose a misspecified AR(1) model as the univariate model for y_t and assume there is no estimation error when computing the residual, the second stage forms

$$\begin{aligned} \begin{bmatrix} \Lambda_1\eta_{t-1} + \epsilon_{1,t} - \phi\epsilon_{1,t-1} \\ x_t \end{bmatrix} &= \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix} \eta_{t-1} + \begin{bmatrix} \epsilon_{1,t} - \phi\epsilon_{1,t-1} \\ \Lambda_2\phi f_{t-1} + \epsilon_{2,t} \end{bmatrix}, \\ &= \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix} \eta_{t-1} + \xi_t. \end{aligned} \tag{23}$$

The third stage simulates signal corresponding to v_t from the second stage DFM. Again, assuming no estimation error we have the i-th simulated prediction error

$$\tilde{v}_t^{(i)} = Z_1(I_p - \kappa\kappa')\tilde{f}_t^{(i)},$$

where $\tilde{f}_t^{(i)}$ is simulated from the Gaussian smoothing density of the factor f_t implied by model (22). Importantly, we see that the ensemble time series remove $\hat{\epsilon}_{1,t}$ and is given by

$$\tilde{y}_t^{(i)} = l\hat{g}_t + \tilde{v}_t^{(i)} = l\sum_{j=1}^{t-1} w_{jt}y_j + Z_1(I_p - \kappa\kappa')\tilde{f}_t^{(i)}. \tag{24}$$

Notice that, in the second stage DFM (22) or (23), the observation error term $(\hat{\epsilon}_{1,t}, \epsilon_{2,t}')'$

becomes colored, resulting in a covariance matrix with a band structure, which is an approximate DFM. We thus follow the two-step method of Doz et al. (2011) to consistently and efficiently estimate the model.

Also from (18), we see that a chosen univariate model fails to account for the part of information related to $Z_1(I_p - \kappa\kappa')f_t$ in v_t and it gives $l\hat{g}_t \perp v_t$. The second stage DFM and the third stage simulation however breaks the orthogonality to recover the part of information about $Z_1(I_p-\kappa\kappa')f_t$ that is lost in the first stage univariate. It is easy to see that the non-zero covariance between $l\hat{g}_t$ and $\tilde{v}_t^{(i)}$ comes from the fact that $\tilde{f}_t^{(i)}$ from the simulation smoother (Durbin and Koopman 2012, Shephard and Pitt 1997) is a function of $v_{1:T}$ and $X_{1:T}$ which are thus correlated with $y_{1:t-1}$.

As argued previously, $\kappa \in \{k \in \mathbb{R}^{p\times q} : k'k = I_q, E(kk') = I_p\}$ is still undetermined. The following shows that the third stage simulation of our proposed method is helpful for extracting as much as information on the dynamics of the unobserved factor f_t from available data. Firstly, define a diagonal matrix

$$\mathfrak{F}_t^{(i)} = diag(\frac{\tilde{f}_{1,t}^{(i)}}{\bar{f}_{1,t}}, \frac{\tilde{f}_{2,t}^{(i)}}{\bar{f}_{2,t}}, ..., \frac{\tilde{f}_{p,t}^{(i)}}{\bar{f}_{p,t}}),$$

where $\bar{f}_t = E(f_t|v_{1:T}, X_{1:T})$, i.e. the smoothed mean of the factor processes implied by the second stage DFM (15), so immediately $E(\mathfrak{F}_t^{(i)}|v_{1:T}, X_{1:T}) = I_p$. We notice that the simulated prediction error equals

$$\tilde{v}_t^{(i)} = Z_1(I_p - \kappa\kappa')\tilde{f}_t^{(i)} = Z_1(I_p - \kappa\kappa')\mathfrak{F}_t^{(i)}\bar{f}_t = Z_1(\mathfrak{F}_t^{(i)} - \dot{\kappa}^{(i)}(\dot{\kappa}^{(i)})')\bar{f}_t,$$

with $\dot{\kappa}^{(i)}$ being any solution to $\dot{\kappa}^{(i)}(\dot{\kappa}^{(i)})' = \kappa\kappa'\mathfrak{F}_t^{(i)}$ which may have complex entries.

Notice also that if we assume that y_1 is fixed, the information set generated by $\{v_{1:T}, X_{1:T}\}$ is the same as the one generated by $\{y_{1:T}, X_{1:T}\}$. The above representation makes clear that simulating signal corresponding to the first stage prediction error v_t from the second stage DFM can be seen as simulating the undetermined κ or $\dot{\kappa}$ from a more "informative" restricted Stiefel manifold

$$\dot{\kappa} \in \{\dot{\kappa} \in \mathbb{C}^{p\times q} : \dot{\kappa}'\dot{\kappa} = I_q, E(\dot{\kappa}^{(i)}(\dot{\kappa}^{(i)})'|y_{1:T}, X_{1:T}) = I_p\}. \tag{25}$$

So we can write the third stage ensemble time series as

$$\tilde{y}_t^{(i)} = l\hat{g}_t + \tilde{v}_t^{(i)} = l\hat{g}_t + Z_1(I_p - \dot{\kappa}^{(i)}(\dot{\kappa}^{(i)})')\bar{f}_t, \tag{26}$$

which is compared with the first stage univariate model

$$y_t = l\hat{g}_t + v_t = l\hat{g}_t + Z_1(I_p - \kappa\kappa')f_t + \epsilon_{1,t} + \hat{\eta}_t,$$

where $\hat{\eta}_t = l(g_t - \sum_{j=1}^{t-1} w_{jt}y_j)$ resulted from estimating the first stage model (16).

A certain chosen univariate model (16) requires a random κ from the space (17), which is intuitively suboptimal, because the weighting matrix κ is also related to the data and original DGP (15). Durbin and Koopman (2012) show that Kalman filter and smoother is the best linear squared error minimizer, so the second and third stage of our method makes use of information in $(y_{1:T}, X_{1:T})$ in a best linear sense. Furthermore, when estimating the third stage ensemble time series (26), we get rid of the noise from $\epsilon_{1,t} + \hat{\eta}_t$ and importantly

$$\mathrm{Var}\big(Z_1(I_p - \dot{\kappa}^{(i)}(\dot{\kappa}^{(i)})')\bar{f}_t\big) < \mathrm{Var}\big(Z_1(I_p - \kappa\kappa')f_t\big). \tag{27}$$

To see this, we notice that both $\mathrm{Var}\big(vec(I_p - \dot{\kappa}^{(i)}(\dot{\kappa}^{(i)})')\big) < \mathrm{Var}\big(vec(I_p - \kappa\kappa')\big)$ and $\mathrm{Var}(\bar{f}_t) < \mathrm{Var}(f_t)$ in the matrix sense, because both left-hand sides operate on a smaller information set. For simplicity we show the latter inequality. By the law of total variance

$$\mathrm{Var}(f_t) - \mathrm{Var}(\bar{f}_t) = \mathrm{Var}\big(E(f_t|y_{1:T}, X_{1:T})\big) + E\big(\mathrm{Var}(f_t|y_{1:T}, X_{1:T})\big) - \mathrm{Var}(\bar{f}_t) = E\big(\mathrm{Var}(f_t|y_{1:T}, X_{1:T})\big),$$

which is a positive definite matrix. Following the formula of variance for product of random variables derived by Goodman (1960) inequality (27) can be shown. It is also possible to derive explicit expression for the variances on both sides of the above inequality if we assume a joint normal distribution for $(vec(\kappa^{(i)})', v_t, X_t')'$.

Therefore the third stage achieves noise reduction. A direct consequence for this is that the signal-to-noise ratio from the chosen univariate model (16) increases. Or to put it differently, the underlying dynamics is strengthened by the consistent estimate $\bar{f}_t$ from the second stage DFM (22) where information of other variables $X_{1:T}$ is taken into account in a

best linear sense. For example, for the simple model (19), the third stage gives

$$\tilde{y}_t^{(i)} = \phi \hat{y}_t + \Lambda_1 \tilde{\eta}_{t-1}^{(i)},$$

after the second stage DFM (23). So the second stage DFM simulates signal with the help of x_t, and it gets rid of the MA(1) part $\epsilon_{1,t} - \phi\epsilon_{1,t-1}$ in (20). As a result, fitting an AR(1) for the third stage ensemble time series mitigate the model misspecification for the first stage univariate model.

The last step of our proposed method averages the forecasts obtained from all ensemble time series. Jensen's inequality shows that indeed averaging over different realizations of $\tilde{v}_t^{(i)}$ or equivalently $\dot{\kappa}^{(i)}$ from the space (25) leads to improved accuracy. It follows that

$$\begin{aligned} &E\big([E_{\dot{\kappa}^{(i)}}(\tilde{y}_t^{(i)}) - y_t]^2\big) = E\big([E_{\dot{\kappa}^{(i)}}(\tilde{v}_t^{(i)}) - v_t]^2\big) \\ =&E\big([E_{\dot{\kappa}^{(i)}}(\tilde{v}_t^{(i)})]^2\big) - 2E\big(E_{\dot{\kappa}^{(i)}}(\tilde{v}_t^{(i)} v_t)\big) + \mathrm{Var}(v_t) \\ \leq&E_{\dot{\kappa}^{(i)}}\big([E(\tilde{v}_t^{(i)})]^2\big) - 2E_{\dot{\kappa}^{(i)}}\big(E(\tilde{v}_t^{(i)})v_t\big) + \mathrm{Var}(v_t) \\ \leq&E_{\dot{\kappa}^{(i)}}\big(E([\tilde{v}_t^{(i)} - v_t]^2)\big), \end{aligned} \tag{28}$$

where $E_{\dot{\kappa}^{(i)}}(.)$ denotes the expectation with respect to the random variables occurring in the process of simulating $\tilde{f}_t^{(i)}$ from the smoothing density implied by the second stage DFM. Though it is not easy to show, another possible effect of averaging is to smooth out the accumulated estimation error when we estimate the first stage univariate model, the second stage DFM using the two-step procedure of Doz et al. (2011) and the third stage ensemble time series.

We simulate 100 data replications from the model (19) with length 1000 and the following parameter values

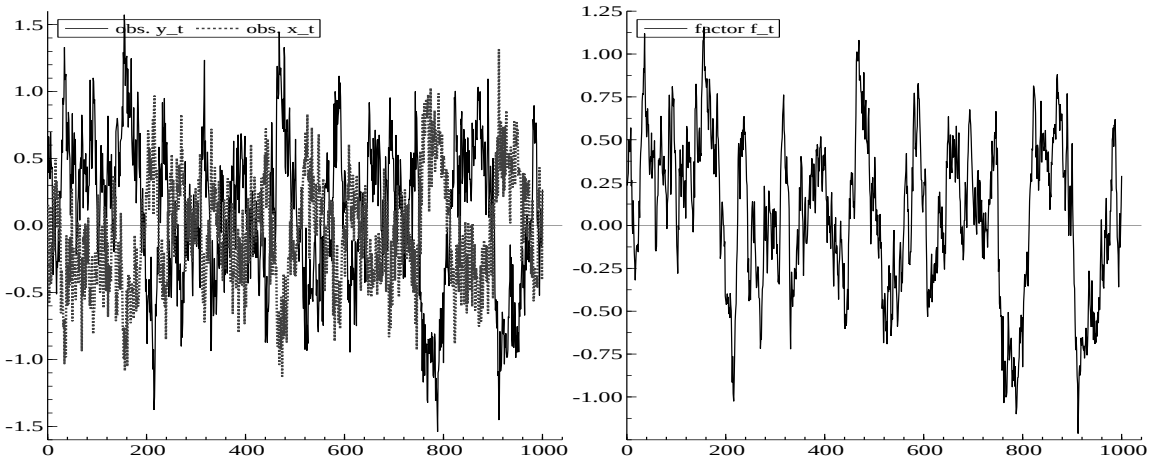
$$\Lambda_1 = 1.2, \quad \Lambda_2 = -0.8, \quad \phi = 0.95, \quad \sigma_{\epsilon_1} = 0.16, \quad \sigma_{\epsilon_2} = 0.18, \quad \sigma_\eta = 0.14.$$

A randomly chosen replication has observation y_t which is the main variable to forecast and x_t as illustrated in the left panel of Figure 13 and the factor f_t in the right panel.

To support previous arguments, we compare h-step ahead forecasting performance with $h = 1, ..., 20$ using different models. The evaluation of forecasts is based on a rolling window

Figure 13

A random simulated dataset from model (19)



Left: y_t and x_t; Right: f_t.

exercise starting with the first half of a replication and each window has length $T = 480$, so for each replication we can collect 500 h-step ahead forecasts. The criterion we use is the following average root mean squared forecast error $ARMSE_h$ for the h-step ahead forecasts

$$ARMSE_h = \frac{1}{M}\sum_{m=1}^{M}\sqrt{\frac{1}{J}\sum_{j=1}^{J}(y_{T+h}^{m,j} - \hat{y}_{T+h}^{m,j})^2},$$

where $y_{T+h}^{m,j}$ is the realized value and $\hat{y}_{T+h}^{m,j}$ the predicted value in the j-th moving window of the m-th replication. In this exercise $M = 100$ and $J = 500$.

We compare six models: the misspecified AR(1) model; AR(1)_DFS which is our proposed method with AR(1) as the first stage univariate model; the correctly specified ARMA(1,1) for the reduced form of y_t, i.e. model (20); ARMA(1,1)_DFS where we choose the correctly specified univariate model; DFM_estPar where we estimate the true DGP (19) (for identification we set $\Lambda_1 = 1$) and make forecasts; DFM_truePar where forecasts are produced via the true DGP with true DGP parameter values. Table 7 shows the $ARMSE_h$ for each model.

The table is visualized in the left and middle panel of Figure 14. DFM_truePar seems to perform the best yielding the smallest $ARMSE_h$ for almost all h considered among the six models, because it is the based on the true DGP without estimation of system parameters. The forecasting accuracy is however worsened if estimation is taken into account. From

Table 7

Forecasting Evaluation of Different Models

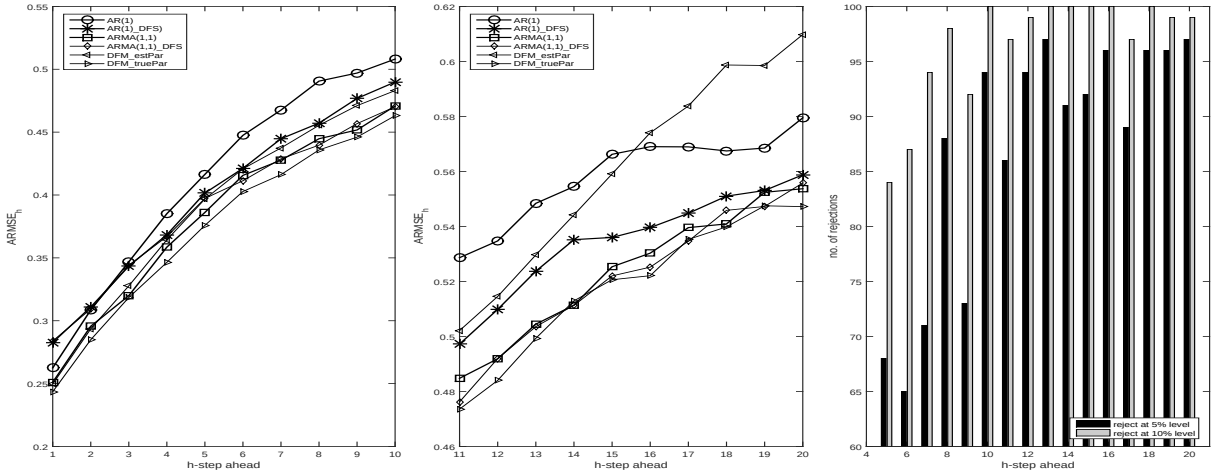
h-step	AR(1)	AR(1)_DFS	ARMA(1,1)	ARMA(1,1)_DFS	DFM_estPar	DFM_truePar
1	0.2598	0.2835	0.2492	0.2832	0.2503	0.2431
2	0.3115	0.3074	0.2907	0.3094	0.2996	0.2845
3	0.3476	0.3456	0.3257	0.3428	0.3336	0.3182
4	0.3840	0.3702	0.3582	0.3646	0.3660	0.3502
5	0.4166	0.4037	0.3861	0.3973	0.3924	0.3807
6	0.4499	0.4237	0.4092	0.4124	0.4200	0.4013
7	0.4680	0.4436	0.4296	0.4276	0.4401	0.4176
8	0.4898	0.4553	0.4408	0.4461	0.4545	0.4391
9	0.5045	0.4714	0.4550	0.4566	0.4742	0.4495
10	0.5156	0.4893	0.4716	0.4674	0.4868	0.4621
11	0.5231	0.4965	0.4753	0.4746	0.4981	0.4726
12	0.5343	0.5069	0.4933	0.4906	0.5155	0.4860
13	0.5449	0.5164	0.5004	0.5064	0.5322	0.4937
14	0.5551	0.5346	0.5136	0.5135	0.5500	0.5085
15	0.5577	0.5371	0.5256	0.5218	0.5595	0.5149
16	0.5656	0.5460	0.5355	0.5273	0.5709	0.5277
17	0.5704	0.5468	0.5383	0.5422	0.5858	0.5335
18	0.5746	0.5483	0.5380	0.5424	0.5936	0.5401
19	0.5757	0.5535	0.5512	0.5486	0.6057	0.5427
20	0.5709	0.5581	0.5585	0.5513	0.6122	0.5493

Reported is the average root mean squared forecast error $ARMSE_h$ for different methods in a rolling window exercise with 100 replications.

$h = 3$, it can be seen that DFM_estPar gives $ARMSE_h$ that is 1.02 times bigger than that given by DFM_truePar, and this difference increases to more than 1.2 when $h = 20$. It is clear that this deterioration of forecasting quality by DFM_estPar goes linearly in h. It is well-known that $h-$step ahead forecasts delivered by dynamic factor models are destined to be contaminated by estimation errors especially for increasing h and model dimensionality. In our simple model, only four parameters need to be estimated and this contamination effect is already manifest. For $h \geq 16$, $ARMSE_h$ given by the correctly specified DFM_estPar is even larger than 0.58 which is never surpassed by the misspecified AR(1) model. This partially explains why when making forecasts, researchers often favor simpler and more parsimonious univariate models (Ashley, 1988) over multivariate models which may seem to follow the underlying DGP closer.

Figure 14

$ARMSE_h$ for different models



Left: $h = 1, ..., 10$; Middle: $h = 11, ..., 20$; Right: Frequency of rejections from DM test for equal predictive accuracy between forecasts given by AR(1)_DFS and by AR(1).

Fitting an AR(1) model to y_t gives the worst forecasts because it misses the MA(1) dynamics for the reduced form model of y_t. As a result, the $ARMSE_h$ given by AR(1) is approximately 1.08 times that given by ARMA(1,1). But if AR(1) is coupled with the proposed dynamic factor simulation procedure, improvements can be achieved. Except for $h = 1$, 2 and 3, AR(1)_DFS delivers $ARMSE_h$ that is clearly smaller than that given by AR(1) and the improvement seems to be more observable when h gets larger. This is can also be found in the right panel of Figure 14 which shows the number of rejections of Diebold-Mariano test based on a squared forecast error loss function. Out of the 100

replications, less than 70 replications show significant improvement from AR(1)_DFS for $h = 5$ and 6 but his number exceeds 85 when $h \geq 8$ at 5% level. At 10% level, significant improvement can be found in almost all replications when $h \geq 8$. This confirms our previous assertions that the second stage DFM (23) helps extract information on the dynamics of f_t using x_t, and the third stage ensemble time series (26) mitigates the first stage AR(1) model misspecification. Therefore although AR(1)_DFS fails to perform better than the correctly specified ARMA(1,1) model is does improve from AR(1) and the resulted $ARMSE_h$ converges to that given by ARMA(1,1) when $h \geq 14$. Lastly, ARMA(1,1)_DFS performs almost identically as ARMA(1,1), which shows that the proposed dynamic factor simulation neither worsens nor improves a first stage univariate model that is already correctly specified.

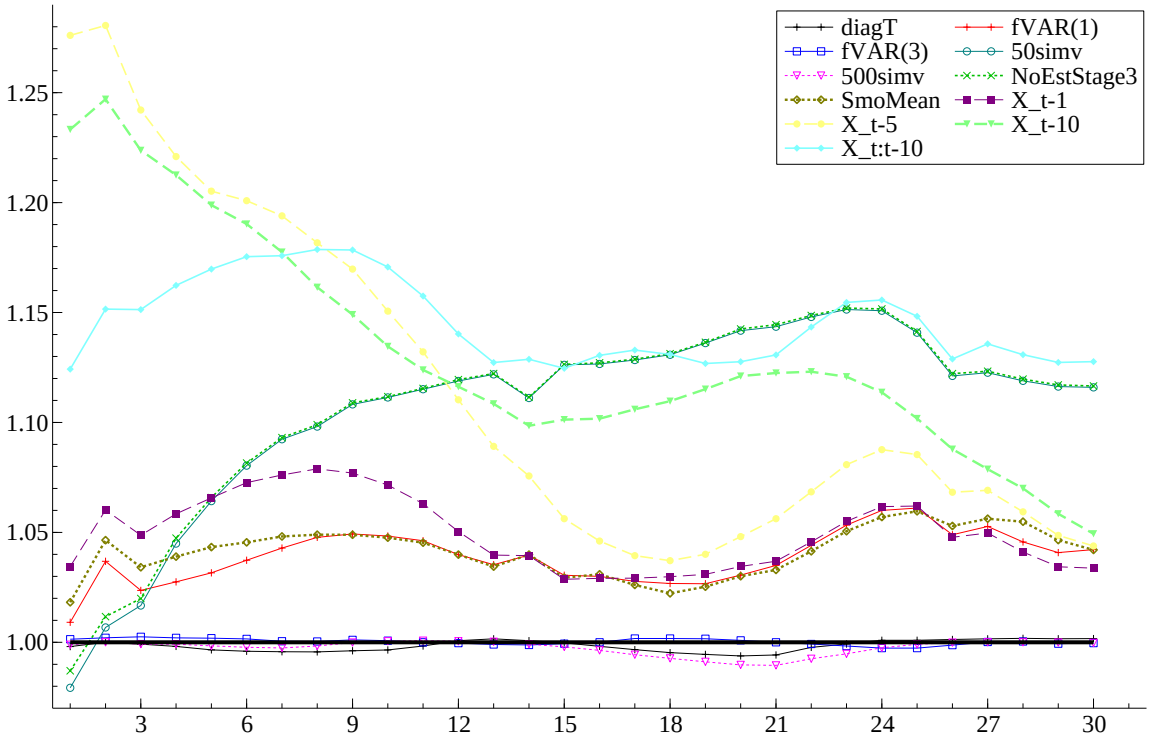
D RMSE among different specifications of UC_DFS

Figure 15 shows the ratio of RMSE of different UC_DFS specifications to that of our base line UC_DFS used in Section 4. "diagT" indicates diagonal transition matrices for the VAR(2) dynamics used for the second stage DFM. Some extent of parsimony is achieved by specifying diagonal transition matrices for the factor dynamics but the results are literally the same as full matrices are used. "fVAR(1)" and "fVAR(3)" indicate a VAR(1) and VAR(3) model is used for the second stage DFM. It can be seen that the RMSE of both specifications is higher than that of the baseline specification for all h, but the difference is smaller than 5% except for $h = 24$ and the HLN_DM test cannot tell statistical difference. "simv" means the number of simulated residual ensembles after the construction of second stage DFM. While "50simv" underperforms the baseline specification (significantly so for $h > 10$) with 200 simulated residual series, increasing the number of simulated residual ensembles to 500 suggest no difference. This means that it is not necessary to have a very large ensemble size for our procedure. "NoEstStage3" means no estimation is carried out for the third stage ensemble Niño3.4 series but we keep the first stage estimated parameters, and "SmoMean" does not simulate the residual ensembles $\tilde{v}_t^{(i)}$ from $p_{\text{dfm}}(v_t^{uc}|X_1, ..., X_T)$ but to use the smoothed mean $E(v_t^{uc}|X_1, ..., X_T)$ as the single ensemble in the third stage. Both of the two specifications perform worse than the baseline specification, highlighting the benefit from third stage estimation and averaging, i.e. reducing estimation errors, which is in line

with our reasoning in (28). "X_t-j" means we do not use the contemporaneous X_t in the second stage DFM, but use the j-th period lagged X_{t-j} in the second step DFM. And "X_t:t-10" is the collection of $(X_t, X_{t-1}, ..., X_{t-10})$. Except for "X_t:t-10", the forecasts from other specifications are clearly inferior to those obtained from our baseline specification, which stresses the use of the most recent explanatory variables X_t to make reliable point forecasts. The fact that one fails to reject the null of equal predicative accuracy between "X_t:t-10" and the baseline specification (and the former seems to perform slightly worse from Figure 15) sheds light on that naive use of older observation such as the S&W and CDFM may not always be a good choice for improving El Niño forecasts. Our choice and specification performs the best in terms of point forecast accuracy except that one can also increase the number of residual ensembles or make the factor transition matrices diagonal.

Figure 15

Ratio of RMSE among different specifications of UC_DFS



The baseline specification has VAR(2) process with full transition matrices for the factors in the second step DFM $(v_t^{uc}, X_t')'$, and 200 ensemble prediction errors $\tilde{v}_t^{(i)}$. All third step ensemble time series $\tilde{y}_t^{(i)}$ for $i = 1, ..., 200$ are estimated.

References

Ashley, R. (1988). On the relative worth of recent macroeconomic forecasts. *International Journal of Forecasting 4*, 363–376.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Barnston, A., M. Chelliah, and S. Goldenberg (1997). Documentation of a highly ENSO-related SST region in the equatorial Pacific. *Atmosphere-Ocean 35*, 367–383.

Bräuning, F. and S. J. Koopman (2014). Forecasting macroeconomic variables using collapsed dynamic factor analysis. *International Journal of Forecasting 30*(3), 572–584.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*, 253–265.

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*, 188–205.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89(3)*, 603–615.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (2nd ed.). Oxford: Oxford University Press.

Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology 8*, 985–987.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics 82*(4), 540–554.

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

Good, S. A., M. J. Martin, and N. A. Rayner (2013). En4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans 118*, 6704–6716.

Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association 55*(292), 708–713.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

Harvey, A. C. and N. Shephard (1993). 10 structural time series models. *Handbook of statistics 11*, 261–302.

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting 13*(2), 281–291.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica 46*, 1251–1271.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph (1996). The ncep/ncar 40-year reanalysis project. *Bulletin of the American Meteorological Society 77*, 437–471.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions – five years of experience. *Journal of Business & Economic Statistics 4*, 25–38.

Petrova, D., S. J. Koopman, J. Ballester, and X. Rodó (2016). Improving the long-lead predictability of El Niño using a novel forecasting scheme based on a dynamic components model. *Climate Dynamics 48*, forthcoming.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*(3), 653–667.

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*, 1167–1179.

Stock, J. H. and M. W. Watson (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics 30*, 481–493.

Tsay, R. S. (2010). *Analysis of Financial Time Series* (3rd ed.). New York: John Wiley and Sons.

Varian, H. R. (1975). *A Bayesian approach to real estate assessment.* North Holland: Amsterdam.