Long-Lead Forecasting of El Niño Southern Oscillation Using Score-Driven Models with Many Predictors *

Janneke van Brummelen[a], Sebastian Jensen[b],

Siem Jan Koopman[a], and Desislava Petrova[c]

[a]Vrije Universiteit Amsterdam, The Netherlands

[b]Aarhus University, Denmark

[c]Barcelona Institute for Global Health (ISGlobal), Barcelona, Catalonia, Spain

Abstract

We consider long-lead forecasting of the El Niño Southern Oscillation climate phenomenon. This is widely seen as a challenging task from a methodological perspective. At the same time, El Niño events are of substantial social importance. We adopt the class of score-driven models to capture key dynamic features of the Niño3.4 index and also include a large set of previously designed predictors. We provide details of our forecasting methodology and assess the forecast accuracy for a range of different model specifications and estimation strategies. We conclude that a model with a good description of the dynamic features in the data is also able to provide accurate forecasts for almost all lead times, up to 18 months ahead. The large set of predictors is shown to support the provision of accurate long-range forecasts for specific lead times.

Keywords: Niño3.4 index time series, ENSO forecasting, general-to-specific, LASSO, observation-driven model, filtering, out-of-sample forecasting, maximum likelihood.

*Corresponding author: Siem Jan Koopman, Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email: s.j.koopman@vu.nl.

1 Introduction

The social relevance of El Niño Southern Oscillation (ENSO) events has become increasingly apparent as a result of their links to multiple climate- and weather-related impacts and disasters worldwide (Sarachik and Cane, 2010), some of which are now exacerbated by climate change (Wilcox et al., 2023). Furthermore, ENSO is the primary source of seasonal climate forecast skills (Ropelewski and Halpert, 1987; Rodó et al., 2006; Sarachik and Cane, 2010; L'Heureux et al., 2020), and its early prediction can help improve outcomes in many sectors of society, such as health, agriculture and energy production, among others. Hence, the provision of accurate ENSO forecasts for long horizons extending to 9, 12, and up to 18 months ahead is of considerable practical relevance.

Institutions that officially provide ENSO forecasts are the International Research Institute for Climate and Society (IRI) Earth Institute[1] and the National Atmospheric and Oceanic Administration with its Climate Prediction Center (NOAA/CPC)[2]. Their operational forecasts are for lead times of up to 6-8 months in advance. At the same time, longer horizons for real-time predictions have been shown to become more relevant and should be explored practically (Petrova et al., 2024 and references therein).

The complexity of forecasting ENSO arises due to the fact that it is an atmosphere-ocean coupled mode of climate variability (Wyrtki, 1975; Philander, 1983), and many dynamical processes interact on various timescales to produce the year-to-year variation in atmospheric circulation of the equatorial Pacific and sea surface temperatures (SST; Capotondi et al., 2015). In addition to that, the stochasticity of high-frequency atmospheric winds in the tropical Pacific can limit the accuracy of predictions at long lead times (Fedorov et al., 2015). There are a wide range of operational ENSO forecasting methods, and many of these rely on complex dynamical models, fewer using statistical formulations designed in the 1980s and 1990s (Barnston et al., 2012). However, recently developed statistical and data-driven methods have proven to be very successful in overcoming the spring predictability barrier in ENSO forecasting, which is due to the low signal-to-noise ratio during the boreal spring (Sarachik and Cane, 2010), and in extending forecast lead times to more than a year in advance, sometimes up to two years ahead (Clarke and Gorder, 2003; Ludescher et al., 2013, 2014; Petrova et al., 2017, 2020; Li et al., 2020; Zhou and Zhang, 2023; Wang

[1] https://iri.columbia.edu/our-expertise/climate/enso/ (last accessed November 9, 2025)

[2] https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/enso_advisory/(last accessed November 9, 2025)

et al., 2023). Most of these methods harness the predictive capacity of the equatorial Pacific upper ocean heat content, where the memory of the coupled system resides, in order to successfully forecast ENSO at these longer time horizons. In particular, Petrova et al. (2017) developed a structural time series model based on dynamic components (trend, seasonal, and time-varying cycles) and designed specific predictor variables that capture zonal wind anomalies and subsequent build-up of subsurface ocean heat at different depths in the western and central tropical Pacific, representing early processes which combined may characterize the origin of El Niño events (Wyrtki, 1985; Ramesh and Murtugudde, 2013; Ballester et al., 2015). In contrast, the majority of operational statistical ENSO models do not integrate subsurface information (Barnston et al., 1999; Ehsan et al., 2024), which is critical for prediction at longer lead times (Wyrtki, 1985; Cane et al., 1986; Jin, 1997; McPhaden, 2003; Ramesh and Murtugudde, 2013; Ballester et al., 2016; Petrova et al., 2017).

More recent empirical models that have documented successful forecasts at long lead times have been based on a variety of strategies and techniques: simple regression including the 20°C isotherm in the equatorial Pacific as a proxy for upper ocean heat content and an Indo-Pacific equatorial wind index (Clarke and Gorder, 2003); network methods and emerging teleconnections between different pairs of tropical grid points selected according to the cross-correlations of climate records at these locations, which manifest strong relationships at long lead times (Ludescher et al., 2013); dynamic factor simulations and applying a three-step forecasting procedure along with relevant measurement of combinations of signals from the same explanatory variables designed by Petrova et al. (2017), which capture persistent dynamics that enables medium- and long-range forecasting (Li et al., 2020); data-driven self-attention-based neural network that relies on the Transformer model (3D Geoformer) applied to predict temperature departures in the three-dimensional upper ocean along with wind stress anomalies in the equatorial Pacific (Zhou and Zhang, 2023); a deep learning model that uses artificial intelligence to define long-lead spatial and temporal dynamical processes related to ENSO-associated heat signals (SST and ocean heat content in the upper 300 meters) across the global ocean, highlighting the relevance of inter-basin interactions for the prediction of ENSO (Wang et al., 2023).

Motivated by the success of recent advances in the long-range statistical prediction of ENSO, in this study we introduce score-driven dynamics in the modeling framework of Petrova et al. (2017). The score-driven model allows the location (or mean) to change over time using stochastic dynamic equations that depend directly on observed variables, leading to straightforward computations (Creal et al., 2013; Harvey, 2013). The location equation can be a composite of the same dynamic components as in the structural time series model, also possibly extended by predictor variables. The updating mechanisms of the components remain the same, but they all share the same prediction error as input. The score-driven model belongs to the class of observation-driven models (Cox, 1981) for which the likelihood function is available in closed form. This key feature implies that parameter estimation and forecasting are straightforward. Furthermore, the score-driven location model can be expressed as a standard autoregressive moving average (ARMA) model. A natural benchmark to verify the relative precision of the predictions from the different model specifications is the autoregressive (AR) model, as its analysis is equally practical. In an extensive empirical study, we first present in-sample estimation results for models with different combinations of components (trend, seasonal, and time-varying cycles) and different selections or weighted combinations of predictor variables. Then we focus on out-of-sample performances of our proposed ENSO modeling framework for forecast horizons ranging from 1 month to 18 months. This empirical study confirms that score-driven models based on the meaningful dynamic component structure of Petrova et al. (2017) can produce accurate forecasts at all forecast horizons, but especially at long-range horizons.

The remainder of the paper is organized as follows. In Section 2, we present the Niño3.4 index data as monthly ENSO measurements, predictor variables, and the score-driven modeling framework including model specification, parameter estimation, and forecasting. The different estimation methods include maximum likelihood, general-to-specific, the least absolute shrinkage and selection operator, and principal components. In Section 3, we discuss the main design of our empirical forecasting study, the details of parameter estimation, both for models with and without predictor variables, the forecast strategy, and the forecast accuracy assessment. We also introduce the benchmark models. In Section 4, we present the empirical in-sample results, including the estimated parameters and the filtered estimates of the dynamic components. In Section 5, we discuss the out-of-sample forecast results and present the relative accuracies of these forecasts by means of model confidence sets. In Section 6, we discuss the main results and provide some conclusions. The Appendix provides further results in tables and figures, and additional details of the score-driven model specification.

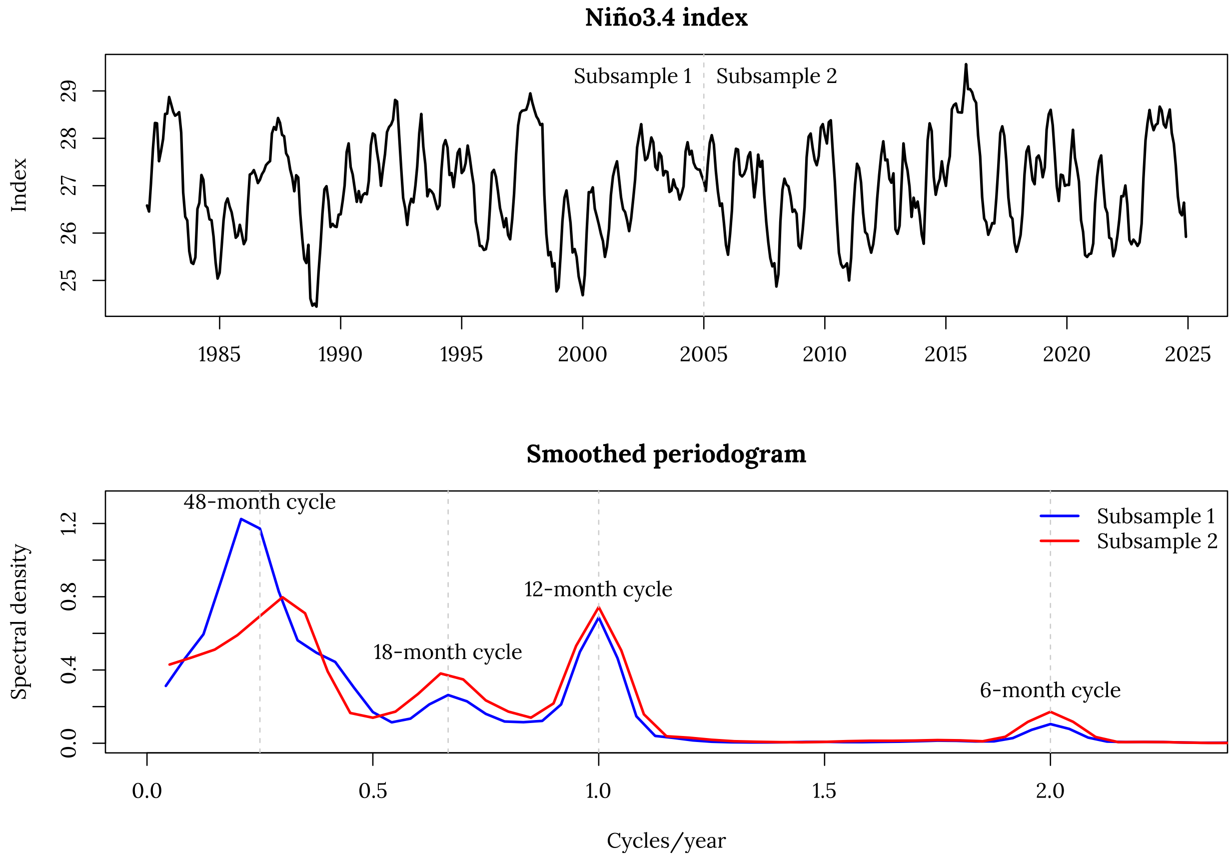

Figure 1: Time series plot of the Niño3.4 index (top) and its smoothed periodogram (bottom).

2 Data, model, estimation and forecasting

2.1 Data description

ENSO events are represented by monthly measurements of the Niño3.4 index (averaged sea surface temperature in the equatorial Pacific region bounded by the box [170°W-120°W]x[5°S-5°N]) from 1982 to 2024, both included, derived from the NOAA OI SST V2 High Resolution Dataset[3] (Huang et al., 2021), which constitutes of $T = 516$ observations. Figure 1 shows a time series plot of the Niño3.4 index together with its smoothed periodogram. To facilitate forecasting, the sample is divided into two subsamples at the beginning of 2005; see Section 3.1. The smoothed periodogram reveals four cycle components, comprising two low-frequency components and two seasonal components corresponding to annual (12-month) and semiannual (6-month) periodici-

[3] https://www.psl.noaa.gov/data/gridded/data.noaa.oisst.v2.highres.html (last accessed November 9, 2025)

ties. The HEGY test (Hylleberg et al., 1990; Table A.1[4]) and the Canova–Hansen test (Canova and Hansen, 1995; Table A.2, top panel) provide no evidence of unit roots at the annual, semiannual, or zero frequencies. The complementary version of the Canova–Hansen test (Table A.2, bottom panel) further rejects seasonal instability.

We use the set of predictor covariates for ENSO forecasting first defined in Petrova et al. (2017), see Table A.3 for a full list and description of the regions from which these variables are extracted. The predictors are specifically designed to anticipate El Niño at long lead times of more than one year ahead of its peak. We use the same dataset employed in the construction of the Niño 3.4 index to compute the sea surface temperature (SST) predictor covariates (see Table A.3). We also add a predictor from the extratropical south Pacific Ocean, the RB SST dipole, which leads El Niño events by 9 months on average (Ballester et al., 2011; Petrova et al., 2024). The subsurface temperature predictor variables extracted from the western and central tropical Pacific Ocean (Table A.3) are calculated using the Hadley Centre EN4.2.2 analyses data with the .g10 bias correction (Good et al., 2013; Gouretski and Reseghetti, 2010; Gouretski and Cheng, 2020). Finally, the zonal wind stress predictors are derived from the NCEP-NCAR reanalysis1 wind data (Kalnay et al., 1996). For further details regarding the physical basis, interpretation, and lag times at which these predictors are relevant, the reader is referred to Petrova et al. (2017).

We utilize a total of 24 different covariate variables (see Table A.3), which may be relevant with delay, possibly stretching from 1 month to 3 years (1-36 months lags). We do not consider all variables and their lags as relevant for Niño3.4 prediction. Table A.3 indicates the relevant lags for each variable, which is based on the previous lead-lag analysis and modeling performed in Petrova et al. (2017). In particular, a distinction is made between short-range (1-5 months), medium-range (6-11 months), long-range (12-18 months) and very long-range (19-36 months).

2.2 Score-driven model formulation

The univariate time series y_t is for the Niño3.4 index in month t and its model is given by

$$y_t = \mu_t + x_t'\beta + \varepsilon_t, \qquad \varepsilon_t \sim IID(0, \sigma^2), \qquad t = 1, \ldots, T, \tag{1}$$

[4]Tables A.1–A.3 are placed in the Appendix.

where μ_t is the time-varying signal variable, x_t is the $k \times 1$ vector of covariates, fixed and known, in month t, β is the corresponding $k \times 1$ vector of fixed regression coefficients, and ε_t is the noise variable with mean zero and variance $\sigma^2 > 0$. In addition, T is the length of the sample and k is the dimension of the vector of covariates x_t.

Motivated by the smoothed periodogram in Figure 1, the time-varying signal μ_t in our model specification (1) is decomposed into three dynamic components: we have $\mu_t = \alpha_t + \gamma_t + \psi_t$ where

(A) location α_t is specified as the random walk process $\alpha_{t+1} = \alpha_t + \theta_\alpha s_t$ with some initial value α_1, some scale θ_α, and zero-mean noise s_t; in case $\theta_\alpha = 0$, α_t reduces to the constant $\alpha = \alpha_1$;

(B) seasonal component γ_t is modeled as a sum of q autoregressive processes of order 2, where the polynomial lag functions have characteristic roots in the complex range with seasonal frequencies;

(C) cyclical component ψ_t is modeled as a sum of r autoregressive processes of order 2, where the characteristic roots are also in the complex range but with admissible ranges for their frequencies corresponding to periods longer than those of the seasonal components.

The motivation for decomposing the signal into trend, seasonal, and cycle components is to enable an accurate description of the dynamic properties of y_t. A description of the signal in terms of these dynamic components can also lead to a better understanding of the key dynamic features of the Niño3.4 index while still providing accurate forecasts. Selecting high values for q and r will improve the in-sample fit, but there is a risk of over-parameterization. In addition, a more extended model may become unstable during the parameter estimation process. We will therefore aim for parsimonious formulations with small q and r. After inspecting the smoothed periodogram of the index Niño3.4 in Figure 1, an obvious choice is $q = r = 2$ with a seasonal periodicity of 6 and 12 months and a cycle periodicity of (approximately) 18 and 48 months. We notice that any seasonal or cycle periodicity can be estimated together with the other parameters of the model.

All dynamic components are driven by the same zero-mean noise sequence s_t, which defines the model (1) as a score-driven model; see Creal et al. (2013) and Harvey (2013). In particular, we define s_t as the score function of the predictive density contribution at time t, as denoted by $p(y_t | y_1, \ldots, y_{t-1}, x_1, \ldots, x_t; \lambda)$, where λ is the parameter vector containing all unknown fixed coefficients in model (1), including β, with respect to the time-varying signal μ_t. More specifically, we define s_t as

$$s_t = S_t \, \frac{\partial \log p(y_t|y_1, \ldots, y_{t-1} x_1, \ldots, x_t; \lambda)}{\partial \mu_t},$$

where S_t is a scaling function such as the inverse of the corresponding information or a function thereof. For all practical purposes, one can also set S_t equal to unity, $S_t = 1$. Using the score for updating the time-varying (sub-)components is intuitive, because the score is the steepest ascent direction to improve the local fit in terms of log likelihood at time t, given the current position of the time-varying parameter μ_t. Thus, the score provides a natural direction for updating the (sub-)components. We notice that this holds regardless of whether the model is correctly specified or not. The conditions for optimality of the score-driven filter are documented in Blasques et al. (2015). Further details of the score-driven model are discussed in Appendix B.

The covariate vector x_t only includes lagged variables. We assume that at time t, the selected variables observed in multiple lagged time periods can potentially have an impact on the ENSO index at time t. Given that many of these selected variables can potentially be good predictors of ENSO for different time periods ahead, we typically assume that, initially, k can be relatively large. However, we always ensure that $k < T$. We adopt three different strategies to select lagged variables in x_t: (a) no predictor covariates (hereafter also regressors) are selected and the model is reduced to $y_t = \mu_t + \varepsilon_t$; (b) selection is based on a *general-to-specific* (G2S) approach where regressors with non-significant coefficient estimates are sequentially removed from the model; (c) a large set of lagged variables is included in x_t and the estimation process relies on the *least absolute shrinkage and selection operation* (LASSO), implying that a smaller set of regressors in x_t is implicitly selected during the estimation process. Given that the set of regressors can be high-dimensional, we also consider dimensionality reduction methods such as *principal component analysis* (PCA). Both G2S and LASSO selection methods can operate on the original high-dimensional data set, but also on the low-dimensional principal components (PCs), computed from the original data set; we indicate the latter option by G2S PCA and LASSO PCA, respectively.

Finally, the model specification in (1) is not complete without an explicit choice for the density function of the disturbance term ε_t, denoted by $p_\varepsilon(\cdot)$. We consider the normal density, that is $\varepsilon_t \sim \text{NID}(0, \sigma^2)$, where σ^2 denotes the disturbance variance. This density leads to a score function that is linear in both the observations and the components.

2.3 Estimation and forecasting

An overview of the econometric methodology for score-driven models is provided by Artemova et al. (2022). The parameter vector λ is estimated by maximum likelihood. The time-varying components α_t, γ_t and ψ_t are unobserved, but for given data sets of y_t and x_t, and for a given value of λ, the filtered values of $\widehat{\alpha}_t$, $\widehat{\gamma}_t$ and $\widehat{\psi}_t$, and thus of $\widehat{\mu}_t$, can be computed recursively for $t = 1, \ldots T$ using the score-driven updating equations. The resulting filtered values are used to evaluate the log-likelihood function, which can then be optimized numerically with respect to λ. For this purpose, some elements in λ may be subject to further transformations to enforce model restrictions such as stationarity, see Appendix B.2. A more detailed discussion on parameter estimation is presented in Appendix B.4.

Given that s_T and therefore $\widehat{\mu}_{T+1}$ is known at time T, one step ahead forecasting is trivial. For larger forecast horizons, we use the assumption that $\mathbb{E}_T[s_{T+k}] = 0$ for $k > 1$, where $\mathbb{E}_T[\cdot]$ denotes the conditional expectation given the information available at time T. Thus, forecasts for more than one step ahead can be obtained by recursively evaluating the filtering equations with s_{T+k} replaced by zero. For regressors, we only consider lag orders that are equal to or greater than the forecast horizon h. The term $x'_{T+h}\beta$ is therefore known at time T and can be used for forecasting.

3 Design of Empirical Forecasting Study

3.1 Forecasting design

The forecasting study is carried out as follows. We split the time series into two periods: 1985-2004 (20 years, first subsample, estimation sample) and 2005-2024 (20 years, second subsample, forecast evaluation sample). In our in-sample analyses, we focus on the first subsample and consider the second subsample only for illustrative purposes. For both subsamples, we effectively also use the data of the three prior years because the covariates in our models may be lagged by up to 36 months. Hence, for the first sample, we use the data from 1982-2004.

We estimate the parameter vectors for various different model specifications and settings, and consider their in-sample performances only for exploratory purposes. The main focus is on an extensive out-of-sample forecast study for a large set of forecast horizons $h \in \{1, 3, 6, 12, 14, 16, 18\}$. We initiate the rolling forecasting process by taking the first subsample, January 1985-December

2004, and estimate the parameter vector for a pre-defined model set. For each model, we produce forecasts for January, March, June, December 2005 ($h = 1, 3, 6, 12$), February, April, June 2006 ($h = 14, 16, 18$). Then, we consider the January 2005 Niño3.4 index variable as a new realization (together with its set of covariates), re-estimate the parameters of the model based on the sample from February 1985 to January 2005, and produce a new set of seven forecasts (for $h = 1, 3, 6, 12, 14, 16, 18$, that is, for February, April, July 2005, January, March, May, July 2006). We repeat this process until we reach the end of the sample (December 2024). In summary, we use a rolling window approach with re-estimation every month. In this way, for each selected model, we generate $20 \times 12 - h + 1 = 241 - h$ forecasts, which are compared with the realizations (from the second subsample), yielding $241 - h$ forecast errors. The rolling window approach is preferred over an expanding window approach, as the former is required for the model confidence set procedure that we discuss below.

3.2 Practicalities of estimation with regressors

Here, we discuss how the methods with regressors have been implemented. As mentioned above, the lag orders considered depend on the forecasting horizon h. In particular, we only consider lag orders h and higher, in order to avoid having to forecast the regressors themselves. Thus, for the models with regressors, we have a different model for every horizon. For any variable, we only consider the lags indicated in Table A.3, see Appendix. Furthermore, the regressors are standardized by subtracting their mean and dividing by their standard deviation to enhance numerical stability and to ensure that they are appropriate for LASSO regression. The mean and standard deviation are determined by the data available at the time of estimation, such that the out-of-sample forecasts are truly out-of-sample. Finally, covariates are detrended in case a significant linear trend is identified at a 5% significance level. This linear trend is again estimated based on the in-sample data.

In total, using all the lag orders indicated in Table A.3, we have a potential of $k = 558$ different covariates, such that $k > T$. However, the methods of G2S and LASSO will lead to much smaller sets of selected covariates, ensuring that k in practice will be such that $k << T$.

Because there are more regressors than observations in our sample, we cannot feasibly apply maximum likelihood estimation if all regressors are included. Instead, we use a tailored G2S approach. We first apply G2S separately to each individual variable with all its relevant lags. Then, G2S is applied once more to the model that is obtained by including all the lags that remained for each variable in the first step. Using a significance level of 1%, we continue to recursively remove the variable-lag combinations that are not significantly different from zero and correspond to the lowest t-value, until all remaining coefficients are significant. To speed up the G2S process, we sometimes remove more than one variable at once, but never more than a quarter of the remaining variables, and also never more than 10 variables.

For the LASSO regression, we use standard implementations. The LASSO penalty parameter is selected based on the in-sample Bayesian Information Criterion (BIC). The LASSO approach is only used for models without score-driven components (i.e. when $\theta_\alpha = 0$, $q = 0$, $r = 0$), because the non-convexity in the parameters of the log likelihood that occurs when score-driven components are included hinders LASSO optimization.

Finally, for models that use PCs instead of the original variables, we construct the PCs separately for each lag between horizon h and 36, by using only the variables relevant to that lag, as indicated in Table A.3. We select the first 6 PCs for every lag order, which explains about 90% of the variation for all of the lags. As for the standardization and detrending, the PCs are constructed solely based on the data that is available at the time of estimation.

3.3 Forecast precision assessment

To assess forecasting precision and determine statistically significant differences across model specifications, we use squared error loss and model confidence sets (MCS), Hansen et al. (2011).

For each forecast horizon h, a MCS is constructed such that it contains the best specification with a given level of confidence, i.e. 90% and 75%. It is done by sequentially eliminating significantly inferior specifications from the initial set based on a sequence of equivalence tests and an elimination rule. The 75% MCS will contain at most as many members as the 90% MCS (and often fewer). We use the bootstrap implementation of the MCS procedure in Hansen et al. (2011), which is simple to use in practice and avoids the need to estimate a high-dimensional covariance matrix. The equivalence test is based on multiple pairwise t-statistics similar to those in Diebold and Mariano (1995) and West (1996), complemented by the natural elimination rule. Further discussion of the MCS validity and the bootstrap implementation is provided in Section 5.2.

3.4 Benchmark models

We compare the forecast accuracy of the score-driven model with that of popular benchmarks: the autoregressive (AR) model with and without G2S-selected regressors, both with and without PCA. We consider a simple AR(1) and an AR(3), where lag order 3 has been selected based on the BIC on the first subsample. To mimic the IRI[5] treatment of statistical ENSO forecasts in their prediction plume (Ehsan et al., 2024), the benchmarks are based on anomaly data, obtained by removing the mean annual cycle from the Niño3.4 index (subtracting the sample mean for each month) before estimation. We compute the mean annual cycle based on the entire sample, i.e. data from 1982 to 2024; see Figure A.1 in Appendix for a plot of the monthly sample means. We emphasize that when estimating the score-driven model in Section 2, we use raw data rather than anomaly data, as the annual cycle is explicitly modeled by seasonal components. In other words, rather than removing seasonality from the data and the model, the score-driven model treats seasonality as an integral part of the model, which is customarily done in operational statistical ENSO forecasting, see Petrova et al. (2020). We have chosen this strategy, because the annual cycle has an impact on the correct evolution of the ENSO cycle phases, the so-called phase-locking of ENSO to the annual cycle (Rasmusson and Carpenter, 1982; Stein et al., 2011). For the benchmark models, we only consider lags of the regressors that are between 36 and the forecasting horizon, and for both benchmarks, we use a standard recursive forecasting approach, just as for the score-driven model. When comparing forecasts across models, we add the mean annual cycle back to the benchmark AR forecasts.

4 In-Sample Results

4.1 Model selection

We consider different configurations of the model described in Section 2, guided by the in-sample fit for the first subsample. Models based on the Student's t distribution turned out to be unsuitable for the data, as the degrees of freedom parameter ν was estimated to be very large, implying that distribution is close to the normal distribution. We therefore restrict attention to the normal distribution. It is well known that the fat-tailed nature of the Student's t distribution causes the

[5]See footnote 1.

score updates to heavily downweight large prediction errors, see e.g. Harvey and Luati (2014), which can yield robustness to outliers. However, the Niño3.4 index is relatively smooth without clear outliers or spikes, and the changes in level are rather persistent. The remaining model choices are as follows.

- Whether to include a stochastic trend ($\theta_\alpha \neq 0$) or a constant ($\theta_\alpha = 0$).
- Whether to include seasonal components ($q > 0$), cyclical components ($r > 0$), both, or neither ($q = r = 0$).
- Whether to include regressors, using either LASSO or G2S, possibly using PCs instead of the original variables.

We include some simple configurations to serve as benchmarks. In particular, we include (1) a model with only a constant and (2) models with two seasonal cycles (i.e. $q = 2$) and either a constant or a trend. For the latter models, we fix the cycle periodicities to 6 and 12 months, to ensure that these are truly seasonal cycles. Two seasonal components are selected, as it suffices to adequately capture any seasonal effects. A similar conclusion was reached by Petrova et al. (2020).

Our most extensive specification for μ_t includes two seasonals $q = 2$, two cycles $r = 2$, and either a constant or a stochastic trend. Here, we do not fix all cycle frequencies of the seasonal components. Instead, we restrict the cycle periodicity of the first cyclical component $\psi_t^{(1)}$ between 13 and 21 months, and that of the second cyclical component $\psi_t^{(2)}$ between 23 and 43 months. As discussed above, these restrictions are imposed to enforce identification and stability, but also to help the model uncover cycles that are in line with expected ENSO dynamical features and the smoothed periodogram in Figure 1. We estimate $\psi_t^{(1)}$ to correspond to a near-annual cycle and $\psi_t^{(2)}$ to correspond to a quasi-biannual/quasi-quadrennial cycle (QB/QQ), similar to Petrova et al. (2017). Such QQ/QB components of ENSO variability have been robustly identified in both surface zonal winds and sea surface temperatures in the equatorial Pacific Ocean by a number of other studies as well (Trenberth, 1976; Rasmusson et al., 1990; Jiang, 1995). The near-annual mode has also been previously linked with the ENSO dynamical evolution and prediction (Jiang, 1995; Jin et al., 2003). For a detailed description of the ocean-atmosphere processes that are associated with these oscillatory ENSO dynamics, we refer the reader to the discussion in Petrova et al. (2017). Increasing the number of cycles r does not improve the in-sample fit, and the estimates of the parameters corresponding to such extra cycles are very insignificant.

We also consider configurations with only regressors and no score-driven components to find out if important periodic variation can be extracted from the regressors and accurate forecasts can be made without any autoregressive dynamics in the model. Finally, we consider combining the elaborate score-driven specification with $q = 2$ and $r = 2$ with regressors.

4.2 In-sample results for models without regressors

Table 1 shows the estimation results for the first subsample, i.e. 1985-2004, for the models without regressors. The models without cycle components have a worse fit than those with cycles, based on the log likelihood. For the model with seasonal components and no additional cyclical components, allowing for a stochastic trend leads to a considerably better fit. For the model with seasonal and cycle components, the fit is slightly worse with a trend. This fit can worsen because the model with the trend technically does not nest the model without the trend. For the trend model, the initialization of the filter $\widehat{\alpha}_1$ is based on the average of the first 6 observations. For the constant model, the initialization α is estimated as a parameter.

The coefficients $\phi_{\gamma,2}^{(i)}$ and $\phi_{\psi,2}^{(j)}$ determine the persistence of the AR(2) processes. Their absolute values correspond to the inverse squared modulus of the roots of the characteristic polynomial of the AR(2) processes, given that we restrict the parameters so that the roots of the AR polynomials are complex. Thus, coefficients close to zero imply less persistence of the corresponding process. Table 1 reveals that the seasonal components tend to have a high persistence. The second cycle component has a moderate persistence. The first cycle component has a very low persistence and the AR coefficients are not significant. On the other hand, the coefficients $\theta_{\psi}^{(j)}$, $j = 1, 2$, are larger than $\theta_{\gamma}^{(i)}$, $i = 1, 2$. Plots of the filtered components for the model with seasonals and cycles without trend are provided in the left column of Figure 2 for the first subsample. The seasonal components are relatively smooth, due to estimates implying a small θ coefficient and a high persistency, whereas the cycles are more irregular.

Table A.4 in Appendix shows the estimation results for the second subsample (2005-2024), for the models without regressors. In the right column of Figure 2, we present the filtered components for the model with constant, seasonals and cycles, hence without a trend, using data from the second sample. The results look very similar to those of the first subsample. One difference is that the AR coefficients of the first cycle component are estimated to be virtually zero, which is why

Table 1: In-sample parameter estimation results for subsample period 1, 1985–1 to 2004–12, and for score-driven models with seasonal ($q = 2$) and cycle ($r = 2$) components, without regressors. Standard errors are reported in italics, LLH is the maximized log likelihood values, AIC and BIC are the Akaike and Schwartz Bayesian information criteria, respectively.

		Seasonals		Seasonals + cycles	
	constant	constant	trend	constant	trend
Constant α	26.941	26.947		26.930	
	0.062	*0.107*		*0.187*	
Trend θ_α			1.135		0.159
			0.066		*0.115*
Variance σ^2	0.908	0.098	0.079	0.071	0.073
	0.084	*0.011*	*0.007*	*0.007*	*0.007*
Seasonal 1 $\gamma_t^{(1)}$ with a fixed periodicity of 6 months					
AR 1 $\phi_{\gamma,1}^{(1)}$		1.000	0.998	0.993	0.993
		n/a	*n/a*	*n/a*	*n/a*
AR 2 $\phi_{\gamma,2}^{(1)}$		-1.000	-0.997	-0.986	-0.986
		0.009	*0.009*	*0.016*	*0.015*
MA $\theta_\gamma^{(1)}$		0.029	0.047	0.073	0.076
		0.015	*0.024*	*0.044*	*0.039*
Seasonal 2 $\gamma_t^{(2)}$ with a fixed periodicity of 12 months					
AR 1 $\phi_{\gamma,1}^{(2)}$		1.115	1.727	1.728	1.729
		n/a	*n/a*	*n/a*	*n/a*
AR 2 $\phi_{\gamma,2}^{(2)}$		-0.414	-0.995	-0.996	-0.997
		0.042	*0.005*	*0.004*	*0.004*
MA $\theta_\gamma^{(2)}$		1.261	0.085	0.077	0.073
		0.068	*0.034*	*0.028*	*0.026*
Cycle 1 $\psi_t^{(1)}$ with implied estim. periodicity				17.063	16.989
AR 1 $\phi_{\psi,1}^{(1)}$				0.363	0.317
				0.316	*0.285*
AR 2 $\phi_{\psi,2}^{(1)}$				-0.038	-0.029
				0.143	*0.171*
MA $\theta_\psi^{(1)}$				0.494	0.402
				0.185	*0.142*
Cycle 2 $\psi_t^{(2)}$ with implied estim. periodicity				29.684	32.513
AR 1 $\phi_{\psi,1}^{(2)}$				1.628	1.645
				0.125	*0.104*
AR 2 $\phi_{\psi,2}^{(2)}$				-0.693	-0.702
				0.120	*0.103*
MA $\theta_\psi^{(2)}$				0.567	0.514
				0.170	*0.128*
LLH	-320.71	-60.72	-34.57	-23.08	-25.08
AIC	645.41	133.43	81.14	70.17	74.16
BIC	652.32	154.16	101.88	111.63	115.62

Figure 2: Data in levels together with filtered components using data of subsample period 1 (left column), 1985–1 to 2004–12, and subsample period 2 (right column), 2005–1 to 2024–12, based on fitted models with constant, seasonals ($q = 2$) and cycles ($r = 2$) with parameter estimates from Table 1 and Table A.4 in the Appendix, respectively. The horizontal solid line in the first row corresponds to the estimated constant α.

they have been removed from the model. The corresponding term $\theta_\psi^{(1)} s_t$ has not been deleted as it is estimated to be significantly different from zero.

4.3 In-sample results for models with regressors

Table 2 summarizes the estimation results for the models with regressors based on the first subsample and horizons 1, 6, 12, and 18. For horizon 1, the log likelihood and sometimes even the

AIC or BIC of the models with only regressors is better than that of the models without regressors in Table 1. From Table 2 we learn that log likelihood values and information criteria worsen as the forecast horizon increases, and the fit of the models with only regressors drop below that of the models without regressors. Across horizons, the AIC and BIC values of the models with both score-driven components and regressors are almost always smaller than their counterparts without regressors in Table 1. These findings are indicative of substantial improvements in the fit when regressors are included.

Typically, the number of selected regressor variables decreases as the horizon increases. For longer horizons, fewer variables are available, and they are likely less informative about the variable of interest due to their higher lag orders. For the models with score-driven components, much fewer regressors are selected than for the models with only regressors. The G2S approach tends to select fewer variables than the LASSO approach, and the log likelihood of the latter is typically higher. Despite the LASSO penalty parameter being based on the BIC, the BIC of the model found by the G2S approach is occasionally better. This may occur because the LASSO in contrast to the G2S approach uses a penalized version of the log likelihood. Lastly, the models based on PCs almost always lead to a worse fit and also worse AIC and BIC values than the models that use the original variables.

Finally, Tables A.5 and A.6[6] show which lags are selected for each variable and horizons 1, 6, 12, and 18, respectively. Table A.7 shows the selected PCs for all lags and horizons. The tables show that not only short lag orders are included. Even for short horizons, high lag orders above 24 are often selected. This stands out in particular for subsurface temperatures, implying that information of more than two years old retains valuable information for forecasting future observations. It highlights the intricate dynamic features that are captured by the subsurface predictor covariates. These covariates are selected to represent the building up and subsurface propagation of heat in the equatorial western and central Pacific Ocean that occurs as a slow process over the course of 2-2.5 years ahead of El Niño events. In this way, warm anomalies captured by the subsurface predictors at these long lead times are highly correlated with the warm surface temperature anomalies at the peak of El Niño (Petrova et al., 2017).

[6]Tables A.5–A.7 are placed in the Appendix.

Table 2: In-sample estimation results for subsample period 1, 1985–1 to 2004–12, and for the score-driven models with only constant and with seasonals ($q = 2$) and cycles ($r = 2$), and both with regressors. Note that selection of lagged regressors in model changes with forecast horizon.

	constant				seasonals + cycles	
	no PCA		PCA		no PCA	PCA
	G2S	LASSO	G2S	LASSO	G2S	G2S
horizon 1						
LLH	138.18	90.94	-45.36	41.77	7.97	-3.75
AIC	-154.37	938.12	136.72	352.45	18.06	41.50
BIC	56.41	2873.10	216.20	1105.71	76.80	100.24
no. selected vars	59	74	21	72	5	5
total vars	558	558	216	216	558	216
horizon 6						
LLH	-43.61	-9.40	-117.45	-70.86	6.91	-3.63
AIC	157.22	948.79	266.90	517.72	20.19	45.26
BIC	278.16	2555.52	322.19	1167.32	78.93	110.92
no. selected vars	33	57	14	51	5	7
total vars	463	463	186	186	463	186
horizon 12						
LLH	-160.03	-4.05	-224.10	-193.93	-4.06	-6.70
AIC	366.05	734.10	470.19	691.85	38.13	43.41
BIC	445.52	1988.38	508.20	1217.06	89.96	95.24
no. selected vars	21	93	9	33	3	3
total vars	361	361	150	150	361	150
horizon 18						
LLH	-207.82	-253.72	-260.31	-267.17	-12.92	-9.67
AIC	449.64	1077.43	536.63	766.35	51.84	55.34
BIC	508.38	2062.20	564.27	1167.16	96.76	117.54
no. selected vars	15	12	6	9	1	6
total vars	283	283	114	114	283	114

5 Out-of-Sample Results

5.1 Forecasting results

Mean squared errors (MSEs) for the forecast evaluation sample 2005-2024 are reported in Table 3. For each forecast horizon h, forecasts in the 90% and 75% MCS are indicated by one and two asterisks, respectively. The difficulty of the forecasting task and forecast uncertainty increases

Table 3: Mean squared forecast errors based on the subsample period 2, 2005–1 to 2024–12. The forecasts in the 90% and 75% model confidence sets are indicated by one and two asterisks, respectively.

	Forecasting horizon h						
	1	3	6	12	14	16	18
Constant	0.971	0.997	1.010	1.035	1.035	1.040	1.037
Regressors							
G2S	0.292	0.778	1.117	1.544	1.520	1.328	1.284
G2S PCA	0.317	0.507	0.748**	1.051*	0.967*	0.881**	0.932**
LASSO	0.205	0.491	0.708**	1.029*	0.923**	0.936*	0.936**
LASSO PCA	0.228	0.521	0.760**	0.946*	0.913**	0.872**	0.925**
Seasonals							
constant	0.112	0.577	0.967	0.983*	0.987*	1.004	1.010
trend	0.092	0.457	1.177	1.615	1.527	1.563	1.688
Seasonals + cycles							
constant	0.082	0.324	0.670**	0.820**	0.815**	0.818**	0.819**
trend	0.083	0.327	0.680**	0.928*	0.923*	0.920*	0.910**
Seasonals + cycles + regressors							
G2S	0.088	0.347	0.699**	0.928*	0.862**	0.851**	0.860**
G2S PCA	0.087	0.328	0.674**	0.844**	0.860**	0.864**	0.847**
Anomaly							
AR(1)	0.082	0.349	0.797**	1.080*	1.014*	0.976	0.974
AR(3)	0.070**	0.280**	0.627**	0.831**	0.812**	0.803**	0.804**
Anomaly + regressors							
ARX(1) G2S	0.079*	0.365	0.904	1.112*	1.046*	0.981**	0.985**
ARX(1) G2S PCA	0.088	0.397	1.041	1.348	1.246	1.086	1.080
ARX(3) G2S	0.077*	0.385	0.809*	0.936**	0.901**	0.844**	0.881**
ARX(3) G2S PCA	0.078	0.309	0.677**	0.924*	0.841**	0.806**	0.816**

with the horizon, translating into less pronounced differences between models and an increasing number of models in the MCS at long horizons. The evolution of the MCS inclusions across horizons indicates that the relative performance between models changes with the forecast horizon.

Across forecast horizons, the specification with seasonals and cycles dominates the specification with only seasonals. The former is always in the 75% MCS for $h \geq 6$ while the latter is rarely in the 90% MCS. This is the pattern that we would expect based on the in-sample results reported in Table 1, where models with only seasonal components have a poorer fit than those with additional cyclical components. It is of less consequence whether the trend component is specified as a constant or a random walk process. This is also in line with the in-sample results and the absence of a zero-frequency unit root. Regarding the former, the signal-to-noise coefficient, θ_α, has an insignificant estimate.

At short forecast horizons one and three months ahead, i.e. $h \in \{1, 3\}$, the specification with seasonals and cycles produces substantially more accurate forecasts than the specification with only regressors. At medium and long forecast horizons at least six months ahead, i.e. $h \in \{6, 12, 14, 16, 18\}$, the specification with only regressors is competitive and included at least in the 90% MCS. This observation is consistent with the fact that the regressors are designed to anticipate El Niño at long lead times, see Section 2.1. Adding regressors to the specification with seasonals and cycles does not lead to a decrease of the mean squared forecast errors, but it does demonstrate comparable results. These findings are not completely in line with the in-sample results, where adding regressors to the models with seasonals and cycles generally led to better information criteria values.

The AR(1), ARX(1), AR(3) and ARX(3) benchmark models are fitted using anomaly data. Their subsequent forecasts are obtained for the data in levels by adding the mean annual cycle as shown in Figure A.1, see Appendix. At short forecast horizons of one and three months, the AR(1) model is among the most accurate models. At medium and long horizons, most of the score-driven specifications produce more accurate forecasts than the AR(1) model. Across horizons, the AR(3) model is often the most accurate and is always included in the 75% MCS. As for the score-driven specification with seasonals and cycles, adding regressors to the AR(1) and AR(3) models generally reduces average forecasting accuracy. However, when analyzing forecasting accuracy as a function of initialization and target month, we note important benefits from adding regressors to the AR(1) and AR(3) models and from the regressors on their own. In the ARX(1) model, PCA has has a negative effect on forecasting accuracy. In the ARX(3) model and the score-driven specification with seasonals, cycles, and regressors, PCA has a positive effect on forecasting accuracy. The same is generally true for the specification with only regressors. This contrasts the in-sample results, where PCA generally deteriorated the in-sample fit. In this out-of-sample study, the information loss caused by PCA is made up for by the lower estimation uncertainty caused by having fewer parameters to estimate.

Figure 3 displays the out-of-sample forecasts for horizons 6, 12, and 18 for four selected models, where the mean annual cycle calculated over the total sample from 1982 to 2024 has been removed from the forecasts. For the longest horizons, the forecasts generated by the models with seasonals and cycles mainly reflect the seasonality, since there is little variation left in the forecasts after the subtraction of the annual cycle. We notice that the estimated persistence of the seasonals is much higher than that of the cyclical components. Therefore, when iterating the filtering equations forward h steps, the seasonals dominate. However, additional cycle components have significant added value, as the models with seasonals only are rarely in the MCS. Figure 3 also shows that many models struggle with the identification of particular El Niño events, especially for the longest horizons. In such cases, the regressors are important for forecasting cold and warm ENSO events. This is further elaborated on in the discussion below.

5.2 Validity checks

Table A.8 in Appendix compares the out-of-sample forecasting performance of our score-driven specifications with stochastic seasonality against specifications with deterministic seasonality or both. The results indicate little difference in forecasting performance across specifications.

In construction of the MCS, we adopt a stationary bootstrap with a block size of 9 and 100,000 replications. The choice of bootstrap parameters is inspired by the empirical inflation forecasting application in Hansen et al. (2011). The results reported in Table 3 are similar when computed with block sizes 6 and 12. A key assumption for the validity of the MCS procedure is that the pairwise forecast loss differentials are stationary across model combinations. The augmented Dickey–Fuller test rejects the unit root null hypothesis for the loss differentials in the vast majority of cases across all model pairs and forecasting horizons.

We have replicated Table 3 for sub-samples that correspond to the El Niño (Table A.9) and La Niña (Table A.10) periods, and they yield similar results. Most notably, the AR(3) model is removed from the 75% MCS at long horizons ($h \in \{16, 18\}$) and even the 90% MCS at the longest horizon ($h = 18$) during El Niño periods. As expected, regressors are important during El Niño, and ARX(1), ARX(3), and models with only regressors are preferred. We extend this discussion with additional results in Section 5.3. We also replicated Table 3 without the benchmark autoregressive models (Table A.11). In general, from these verifications we may conclude that the

MCSs in Table 3 are valid and robust to different model and sample selections.

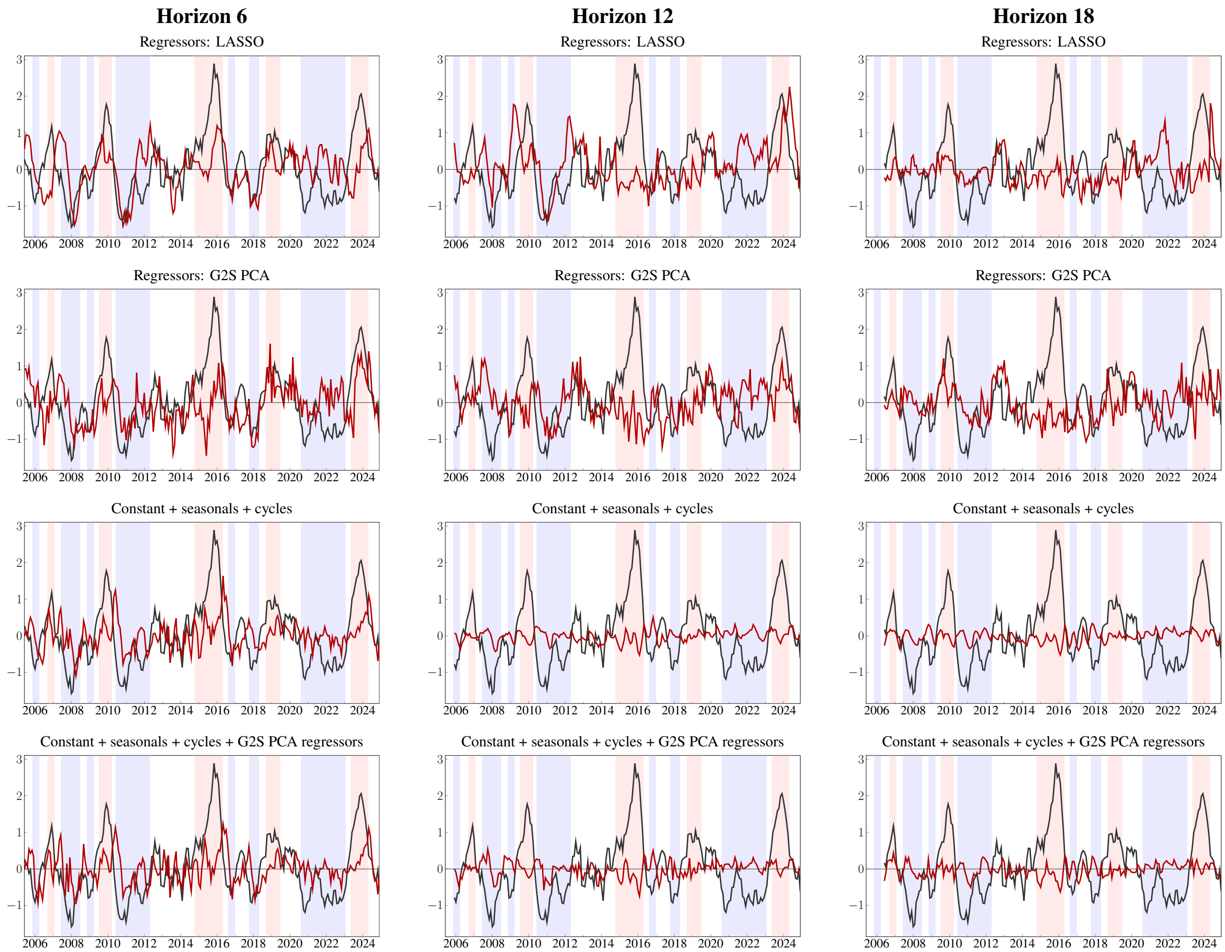

Figure 3: Out-of-sample forecasts (in red) for a selection of models, as indicated on top of each graph, together with the realized Niño3.4 index time series (in black), with the mean annual cycle subtracted from both, for horizons 6, 12, and 18. Red and blue shadings indicate El Niño and La Niña periods, respectively.

5.3 Spring Predictability Barrier

Due to the cyclical nature of ENSO and the need for updated monthly forecasts in real-time, exemplified by the IRI[7] prediction plume, it is important to understand how the accuracy of the forecast varies with the initialization and the target month. In Figure 4, we present heat maps of the mean squared forecast error (MSE) as a function of the initialization month (first axis) and the forecast horizon (second axis), for the four models in Figure 3 and for the AR(1), ARX(1), AR(3), and ARX(3) benchmarks. At the six month horizon, the MSE tends to increase with the onset of

[7] See footnote 1.

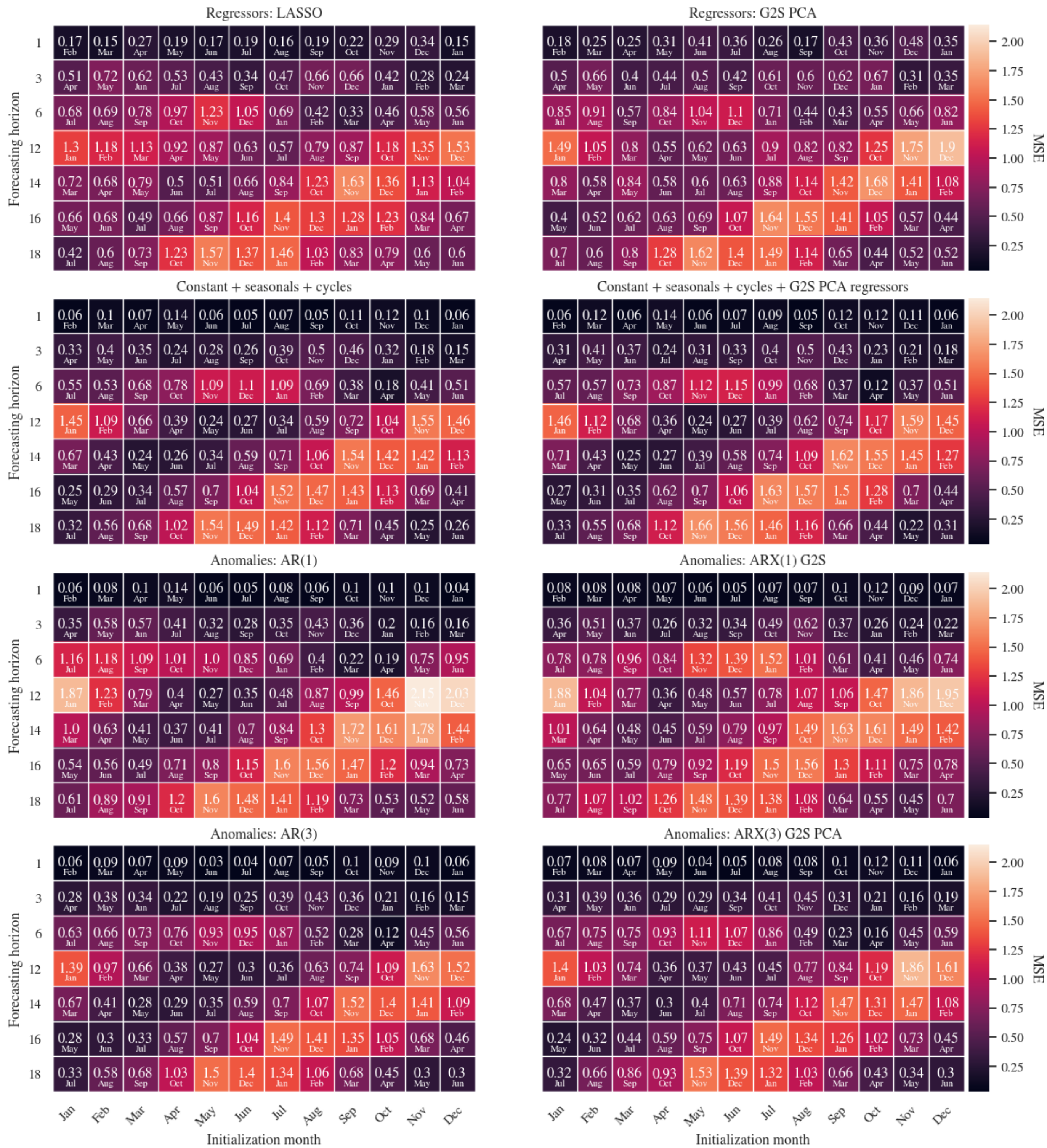

Figure 4: Heatmaps of mean squared forecast errors as a function of initialization month (first axis) and forecasting horizon (second axis), for eight different model specifications which are indicated on top of each heatmap. The target month is included below the MSE value for reference.

boreal spring (March) and does not recover until after spring. This phenomenon is often referred to as the spring predictability barrier and coincides with the usual transition phase of ENSO (Ehsan et al., 2024). The AR(1) and ARX(1) models are particularly vulnerable to this barrier. The MSE

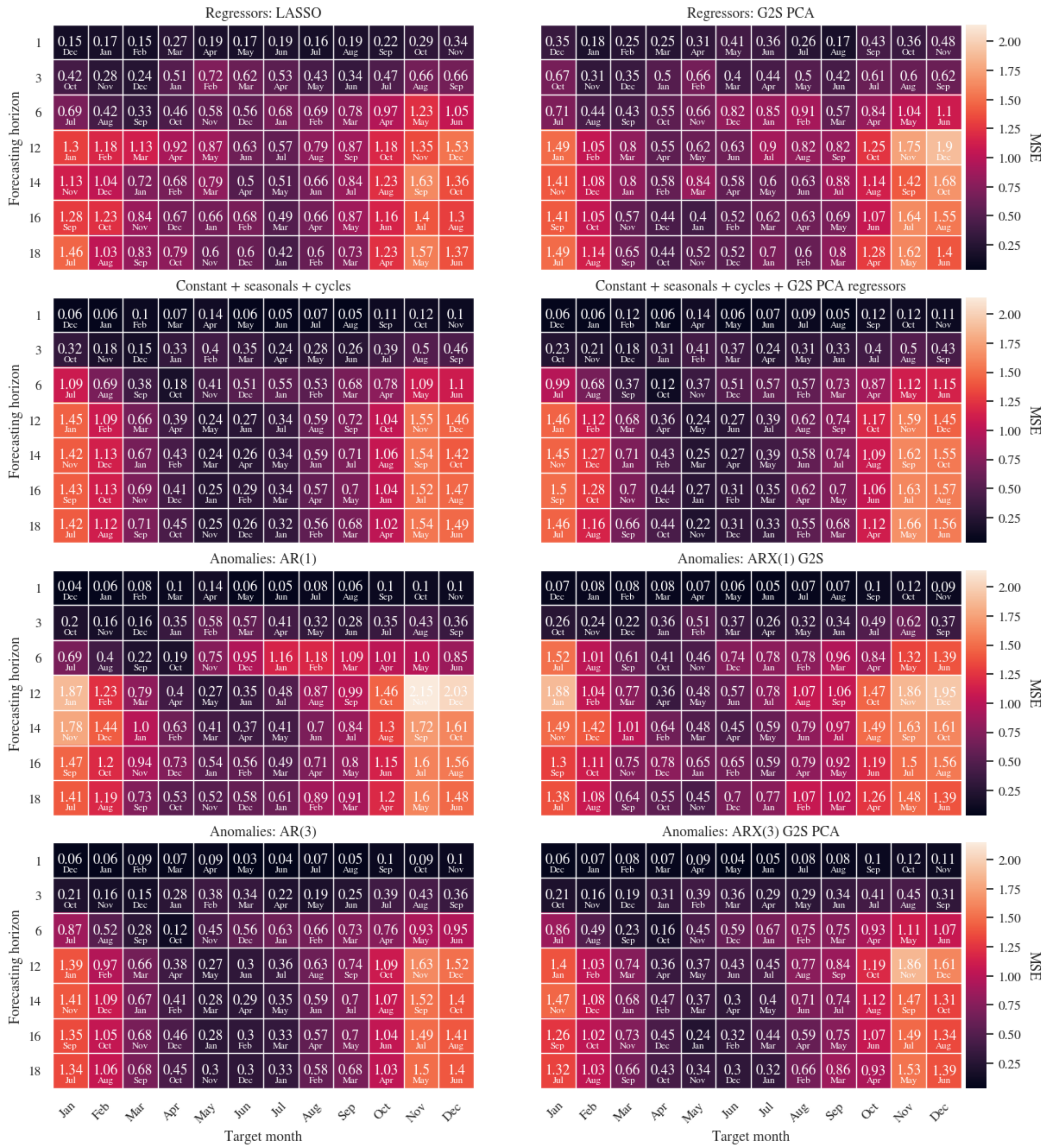

Figure 5: Heatmaps of mean squared forecast errors as a function of target month (first axis) and forecasting horizon (second axis), for eight different model specifications which are indicated on top of each heatmap. The initialization month is included below the MSE value for reference.

for the AR(1) model increases sharply already in January, whereas the ARX(1) model exhibits persistently elevated MSE values through July. In contrast, the spring predictability barrier is less pronounced for the remaining models.

In Figure 5, we present heat maps of the MSE as a function of the target month (first axis) and the forecast horizon (second axis) to supplement the analysis. ENSO typically reaches peak intensity during the boreal winter months (December–February), making this period particularly relevant for forecast evaluation. At long horizons, no single model consistently dominates during boreal winter, although the AR(1) model frequently delivers the weakest performance. The inclusion of regressors tends to improve forecast accuracy at long horizons during boreal winter, especially in the AR(1) and AR(3) models (bottom two rows of Figure 5). In addition, models relying exclusively on regressors remain highly competitive in this setting, and the model based solely on LASSO-selected regressors is often among the strongest performers. These findings are consistent with the construction of the regressor set, which is specifically designed to anticipate El Niño events at long lead times (see Section 2.1). Finally, it is also consistent with the results presented in Figure 3, where models with regressors are found to predict strong ENSO events accurately at the 12 month lead time, and even some models predict them at the 18 month lead time.

5.4 Real-time out-of-sample forecasts

We constructed real-time out-of-sample forecasts from October 2024 onward, which can be evaluated as time progresses. Although December 2024 is the last month for which we have an observation of the Niño3.4 index, we only have observations for all explanatory variables up to and including October 2024. For the parameter estimates, we again use an estimation window length of 20 years, i.e. observations from November 2004 to October 2024. Figure 6 presents out-of-sample real-time forecasts for all selected models, for horizons 1, 3, 6, 12, 14, 16 and 18 months, with the mean annual cycle subtracted, such that the resulting values can be evaluated as forecasts of the anomaly series. The numerical values of the forecasts can be found in Table A.12, see Appendix. The models with only regressors show a wider range of forecast values than the models with cycle and seasonal components and regressors. Especially for horizons $h = 14$ and $h = 16$, the G2S, G2S PCA, and LASSO models without components indicate that moderate to strong La Niña could develop in the winter of 2025-2026. The forecasts from the other models rarely come below the last observed value, and instead these models predict more of an upward movement in the index over the coming 18 months. This finding points towards a transition to more neutral conditions in the equatorial Pacific. Given that an El Niño event took place in 2023-2024, it is unlikely that there will be another event within the next 18 months from October 2024, which is reflected by these forecasts.

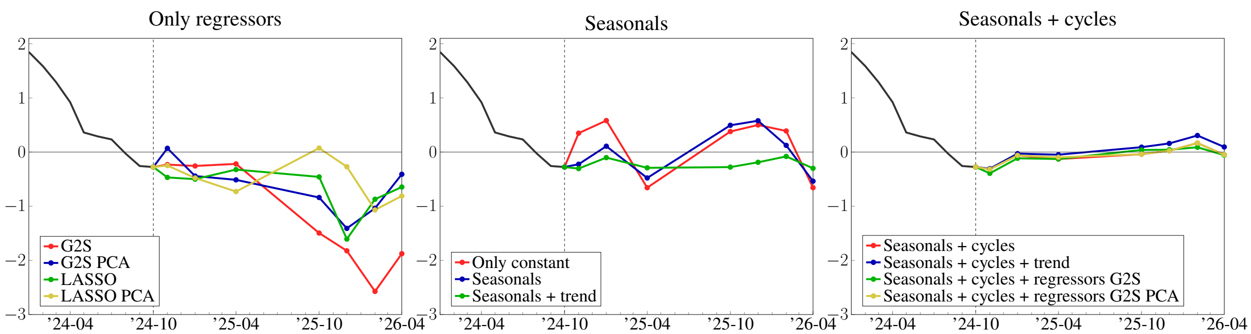

Figure 6: Real-time out-of-sample forecasts made on October 2024 for horizons $h = 1, 3, 6, 12, 14, 16, 18$, with per-month sample mean subtracted, leading to forecasted anomalies.

6 Discussion and Conclusion

We have carried out a comprehensive forecast study for the Niño3.4 index, for the short, medium, and long-lead forecast horizons. The main forecast methodology has relied on a statistical dynamic model that is extended with a large set of predictors. The candidate predictors for the model are selected differently for each forecast horizon. The dynamic features of the trend, seasonal, and cyclical effects are modeled within a score-driven framework, which leads to basic filtering operations and straightforward log likelihood evaluations. Parameter estimation is carried out by the method of classical maximum likelihood. The forecast functions can then be easily obtained. Forecast precision is compared between different model specifications based on the mean squared forecast error statistic and the model confidence set procedure.

Among the score-driven models and models with only regressors, the specification with seasonal and cycle components performs best overall when considering the mean squared error statistic, especially at the shorter forecast horizons (Table 3). However, the specification with only regressors delivers statistically indistinguishable performance at long lead time horizons of 14, 16, and 18 months. This is in line with the design of the covariates to capture dynamic properties of the ENSO phenomenon at very long lead times. We consider as benchmarks AR(1) and AR(3) models with and without regressors based on anomaly time series, obtained by removing the mean annual cycle from the Niño3.4 index before estimation. We find that the AR(1) model is only among the most accurate models at short horizons, while the AR(3) model is often the most accurate model across horizons.

The forecasts presented in Figure 3 reveal a more nuanced picture. At the short lead time horizon of 6 months, the models containing only regressors and the LASSO and G2S PCA configurations accurately predict the major El Niño and La Niña events. The models with seasonals and cycles or seasonals, cycles, and regressors also predict the majority of the large ENSO events but are somewhat underestimated and sometimes with shifted peaks. It is of concern that the large 2015-2016 El Niño is not predicted by the components models. This is possibly due to the low-frequency ENSO dynamics, such as 5-7-year periodicity or decadal variability, which are not represented in these schemes, and they have been shown to be important for the prediction of particular ENSO events (see the discussion in Petrova et al. (2017), especially regarding the other very strong El Niño of 1997-1998.)

At the medium-range forecast horizon of 12 months (middle column of Figure 3), the models with only regressors are again the ones that predict almost all the El Niño events, albeit with shifted peaks (the 2006-2007, 2009-2010 and 2023-2024 El Niños), but no model configuration anticipates the 2015-2016 El Niño. The models with only regressors and LASSO and with G2S PCA configurations also generally predict the La Niña events, although none of the models predicted the triple-dip cold events between 2020-2023, and the regressors with G2S PCA model configuration also missed the 2007-2008 La Niña.

For the long-lead time horizon of 18 months again the models with only regressors perform well at predicting some of the major ENSO events (the 2009-2010 and the 2023-2024 El Niño, and the 2010-2011, 2017-2018 La Niña), but the amplitudes are clearly underestimated, in some cases substantially. Accurately predicting the amplitudes of the events at such long lead times, however, is a challenge for nearly all models, as previously discussed in the literature (Barnston et al., 2012; Petrova et al., 2020). For example, both the operational dynamical and statistical models consistently underestimated the peak intensity of the 2015-2016 El Niño, and only predicted that the SST anomaly would exceed +2°C in July and August of 2015, i.e. a lead time of 3-4 months (L'Heureux et al., 2017). The operational statistical models in fact failed to foresee the actual amplitude, which is consistent with previous ENSO forecasts, as the latest observational information is not readily integrated into these schemes (Barnston et al., 2012).

Finally, the models with only regressors are also the ones that generally predict more accurately the ongoing La Niña event in the winter of 2024-2025, and most of them also indicate that another La Niña might occur in the winter of 2025-2026 (Figure 6 and Table A.12). The other models tend to suggest that the conditions will be neutral in the Niño3.4 region. Interestingly, at the time of writing, the latest official ENSO forecast of the CPC using data up to January of 2025 also points to almost equal probabilities for neutral or La Niña conditions, with very small chances of an El Niño[8].

Given these empirical findings, we may conclude as follows. The regression variables that are specifically selected to account for anomalous atmospheric and oceanic subsurface processes in the tropical Pacific, occurring before individual El Niño events, are particularly relevant for the prediction of cold and warm ENSO events. Moreover, the model configurations with regressors included alleviate the spring predictability barrier and can provide useful information about ENSO events, more than 9 months in advance. However, the inability to predict the strong 2015-2016 El Niño implies that more work should be done in the future to model the potentially nonlinear ENSO dynamics that arises due to the superposition and interaction of low- and high-frequency oscillations involved in the physics of the coupled ocean-atmosphere system.

Acknowledgments

We thank the Editor, Associate Editors, and Referees for their insightful and supportive comments, which have led to a much improved paper. All remaining errors are ours alone. This work was supported by the Independent Research Fund Denmark (grant 4259-00012B to Sebastian Jensen).

References

Artemova, M., Blasques, F., van Brummelen, J., and Koopman, S. J. (2022). Score-driven models: Methodology and Theory. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

[8]https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/enso_advisory/ensodisc.pdf (last accessed November 9, 2025)

Ballester, J., Bordoni, S., Petrova, D., and Rodo, X. (2015). On the dynamical mechanism explaining the western pacific subsurface temperature buildup leading to ENSO events. *Geophysical Research Letter*, 42:2961–2967.

Ballester, J., Petrova, D., Bordoni, S., Cash, B., Garcia-Diez, M., and Rodo, X. (2016). Sensitivity of El Niño intensity and timing to preceding subsurface heat magnitude. *Scientific Reports*, 6.1:36344–.

Ballester, J., Rodriguez-Arias, M. A., and Rodo, X. (2011). A new extratropical tracer describing the role of the western Pacific in the onset of El Niño: Implications for ENSO understanding and forecasting. *Journal of Climate*, 24:1425–1437.

Barnston, A., Glantz, M. H., and He, Y. (1999). Predictive skill of statistical and dynamical climate models in sst forecasts during the 1997-98 El Niño episode and the 1998 La Niña onset. *Bulletin of the American Meteorological Society*, 80:217–243.

Barnston, A., Tippett, M., L'Heureux, M., Li, S., and DeWitt, D. (2012). Skill of real-time seasonal ENSO model predictions during 2002-11. Is our capability increasing? *American Meteorological Society*, 93:631–651.

Blasques, F., Gorgi, P., Koopman, S. J., and Wintenberger, O. (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics*, 12(1):1019–1052.

Blasques, F., Koopman, S. J., and Lucas, A. (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika*, 102(2):325–343.

Blasques, F., van Brummelen, J., Gorgi, P., and Koopman, S. J. (2024). Maximum likelihood estimation for non-stationary location models with mixture of normal distributions. *Journal of Econometrics*, 238(1):105575.

Blasques, F., van Brummelen, J., Koopman, S. J., and Lucas, A. (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics*, 227(2):325–346.

Caivano, M., Harvey, A. C., and Luati, A. (2016). Robust time series models with trend and seasonal components. *SERIEs*, 7(1):99–120.

Cane, M., Zebiak, S., and Dolan, S. (1986). Experimental forecasts of El Niño. *Nature*, 321:827–832.

Canova, F. and Hansen, B. E. (1995). Are seasonal patterns constant over time? a test for seasonal stability. *Journal of Business Economic Statistics*, 13(3):237–252.

Capotondi, A., Wittenberg, A. T., Newman, M., Di Lorenzo, E., Yu, J.-Y., Braconnot, P., Cole, J., Dewitte, B., Giese, B., Guilyardi, E., et al. (2015). Understanding ENSO diversity. *Bull. Amer. Meteor. Soc.*, 96:921–938.

Clarke, A. J. and Gorder, S. V. (2003). Improving El Niño prediction using a space-time integration of Indo-Pacific winds and equatorial Pacific upper ocean heat content. *Geophysical Research Letters*, 30.7:–.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, 8(2):93–115.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Ehsan, M. A., L'Heureux, M. L., Tippett, M. K., Robertson, A. W., and Turmelle, J. (2024). Real-time enso forecast skill evaluated over the last two decades, with focus on the onset of enso events. *npj Climate and Atmospheric Science*, 7:301.

Fedorov, A. V., Hu, S., Lengaigne, M., and Guilyardi, E. (2015). The impact of westerly wind bursts and ocean initial state on the development and diversity of El Niño events. *Climate Dyn.*, 44:1381–1401.

Good, S. A., Martin, M. J., and Rayner, N. A. (2013). EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118:6704–6716.

Gouretski, V. and Cheng, L. (2020). Correction for systematic errors in the global dataset of temperature profiles from mechanical bathythermographs. *Journal of Atmospheric and Oceanic Technology*, 37.5:841–855.

Gouretski, V. and Reseghetti, F. (2010). On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database. *Deep Sea Research Part I: Oceanographic Research Papers*, 57.6:812–833.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, volume 52. Cambridge University Press.

Harvey, A. C. and Luati, A. (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, 109(507):1112–1122.

Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., Smith, T., and Zhang, H.-M. (2021). Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) version 2.1. *Journal of Climate*, 34:2923–2939.

Hylleberg, S., Engle, R., Granger, C., and Yoo, B. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, 44(1):215–238.

Jiang, Ning, J. D. N. M. G. (1995). Quasi-quadrennial and quasi-biennial variability in the equatorial Pacific. *Climate Dynamics*, 12:101–112.

Jin, F. (1997). An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model. *J. Atmo. Sci.*, 54:811–829.

Jin, F., Kug, J., An, S., and Kang, I. (2003). A near-annual coupled ocean-atmosphere mode in the equatorial Pacific Ocean. *Geophysical Research Letters*, 30.2:2923–2939.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.

(1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77:437–471.

L'Heureux, M. L., Levine, A. F., Newman, M., Ganter, C., Luo, J.-J., Tippett, M. K., and Stockdale, T. N. (2020). ENSO prediction. *El Niño Southern Oscillation in a changing climate*, pages 227–246.

Li, M., Koopman, S. J., Lit, R., and Petrova, D. (2020). Long-term forecasting of El Niño events via dynamic factor simulations. *Journal of Econometrics*, 214.1:46–66.

Ludescher, J., Gozolchiani, A., Bogachev, M. I., Bunde, A., Havlin, S., and Schellnhuber, H. J. (2013). Improved El Niño forecasting by cooperativity detection. *Proc. Natl. Acad. Sci.*, 110.29:11742–11745.

Ludescher, J., Gozolchiani, A., Bogachev, M. I., Bunde, A., Havlin, S., and Schellnhuber, H. J. (2014). Very early warning of next El Niño. *Proc. Natl. Acad. Sci.*, 111.6:2064–2066.

L'Heureux, M. L., Takahashi, K., Watkins, A. B., Barnston, A. G., Becker, E. J., and Tom E. Di Liberto, F. G. e. a. (2017). Observing and predicting the 2015/16 El Niño. *Bulletin of the American Meteorological Society*, 98:1363–1382.

McPhaden, M. (2003). Tropical Pacific Ocean heat content variations and ENSO persistence barriers. *Geophys. Res. Lett.*, 30:1480–.

Petrova, D., Ballester, J., Koopman, S. J., and Rodó, X. (2020). Multiyear statistical prediction of ENSO enhanced by the tropical Pacific observing system. *Journal of Climate*, 33.1:163–174.

Petrova, D., Ballester, J., Koopman, S. J., and Rodó, X. (2024). The 2023/24 El Niño and the feasibility of long-lead ENSO forecasting. *Bull. Amer. Meteor. Soc*, 105.10:E1915–E1928.

Petrova, D., Koopman, S. J., Ballester, J., and Rodó, X. (2017). Improving the long-lead predictability of El Niño using a novel forecasting scheme based on a dynamic components model. *Climate Dynamics*, 48.3:1249–1276.

Philander, S. G. H. (1983). El Niño Southern Oscillation phenomena. *Nature*, 302.5906:295–301.

Ramesh, N. and Murtugudde, R. (2013). All flavours of El Niño have similar early subsurface origins. *Nature Climate Change*, 3:42–46.

Rasmusson, E. and Carpenter, T. (1982). Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Weather Rev.*, 110:354–384.

Rasmusson, E. M., Wang, X., and Ropelewski, C. F. (1990). The biennial component of ENSO variability. *Journal of Marine Systems*, 1:71–96.

Rodó, X., Rodriguez-Arias, M., and Ballester, J. (2006). The role of ENSO in fostering teleconnection patterns between the tropical north Atlantic and the western Mediterranean basin. *CLIVAR Exchanges*, 11:26–27.

Ropelewski, C. and Halpert, M. (1987). Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Monthly Weather Review*, 115.8:1606–1626.

Sarachik, E. and Cane, M. (2010). *The El Niño Southern Oscillation Phenomenon*. Cambridge University Press.

Stein, K., Timmermann, A., and Schneider, N. (2011). Phase synchronization of El Niño–southern oscillation with the annual cycle. *Phys. Rev. Lett.*, 107:–.

Trenberth, K. E. (1976). Spatial and temporal variations of the Southern Oscillation. *Quarterly Journal of the Royal Meteorological Society*, 102.433:639–653.

Wang, H., Hu, S., and Li, X. (2023). An interpretable deep learning ENSO forecasting model. *Ocean-Land-Atmosphere Research*, 2:0012.

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

Wilcox, P. S., Mudelsee, M., Spötl, C., and Edwards, R. L. (2023). Solar forcing of ENSO on century timescales. *Geophysical Research Letters*, 50(20):e2023GL105201.

Wyrtki, K. (1975). El Niño - the dynamic response of the equatorial Pacific Ocean to atmospheric forcing. *Journal of Physical Oceanography*, 5:572–584.

Wyrtki, K. (1985). Water displacements in the Pacific and the genesis of El Niño cycles. *Journal of Geophysical Research*, 90:7129–7132.

Zhou, L. and Zhang, R.-H. (2023). A self-attention–based neural network for three-dimensional multivariate modeling and its skillful ENSO predictions. *Sci. Adv.*, 9.10:eadf2827.

Appendix

A Supplementary Tables and Figures

Table A.1: HEGY test for seasonal unit roots at frequencies modeled. Lag order is selected by AIC separately for each deterministic specification; p-values are based on response-surface regressions.

Test	Frequency	Cycles/year	Interpretation	Statistic	p-value
Regression with constant only (AIC lag order = 3)					
t_1	0	0	Zero-frequency root	-4.662	< 0.001
$F_{3:4}$	$\pi/6$	1	Annual cycle	4.988	0.008
$F_{5:6}$	$\pi/3$	2	Semiannual cycle	4.176	0.018
Regression with constant + seasonal dummies (AIC lag order = 2)					
t_1	0	0	Zero-frequency root	-4.232	< 0.001
$F_{3:4}$	$\pi/6$	1	Annual cycle	21.316	< 0.001
$F_{5:6}$	$\pi/3$	2	Semiannual cycle	13.480	< 0.001

Table A.2: Canova-Hansen test for seasonal stability. Regressions include a lagged dependent variable; covariance matrix truncation at lag 12; p-values are based on response-surface regressions.

Test of no unit roots at seasonal frequencies				
Frequency	Cycles/year	Interpretation	Statistic	p-value
$\pi/6$	1	Annual cycle	0.190	0.769
$\pi/3$	2	Semiannual cycle	0.259	0.576

Test of stable seasonal intercepts		
Month	Statistic	p-value
Jan	0.104	0.614
Feb	0.043	0.956
Mar	0.094	0.662
Apr	0.066	0.832
May	0.052	0.914
Jun	0.087	0.701
Jul	0.164	0.381
Aug	0.159	0.396
Sep	0.119	0.542
Oct	0.097	0.647
Nov	0.130	0.497
Dec	0.374	0.088
Joint	1.586	0.474

Note: Covariance matrix truncation at lag 4 leads to similar results.

Table A.3: Lags of regression variables included in models based on previous lead-lag analysis and estimations performed in Petrova et al. (2017). The time series of these regression variables are extracted by averaging over the regions in the tropical Pacific provided in the table. More details about the selection process are given in Petrova et al. (2017).

	short-range (1-5 months)	medium-range (6-11 months)	long-range (12-18 months)	very long-range (19-36 months)
Sea surface temp.				
RB [260°E-280°E]x[65°S-50°S]		✓		
WPAC [140°E-160°E]x[5°S-5°N]	✓		✓	
WPAC2 [140°E-180°E]x[10°S-5°N]	✓		✓	
WPAC3 [120°E-170°E]x[10°S-5°N]	✓		✓	
WPAC4 [140°E-160°E]x[10°S-0°]	✓		✓	
Subsurface temp.				
50m [120°E-170°E]x[10°S-5°N]	✓		✓	✓
100m "cold" [140°E-210°E]x[5°N-10°N]		✓		✓
100m Region 1 [120°E-140°E]x[10°S-5°N]	✓	✓		✓
150m Region 1 [120°E-140°E]x[10°S-5°N]	✓	✓		✓
200m Region 1 [120°E-140°E]x[10°S-7°N]	✓	✓		✓
250m Region 1 [120°E-140°E]x[7°S-7°N]	✓	✓	✓	✓
300 Region 1 [120°E-140°E]x[7°S-7°N]	✓	✓		✓
400m Region 1 [120°E-140°E]x[5°S-5°N]	✓		✓	✓
500m Region 1 [120°E-140°E]x[5°S-5°N]	✓	✓		✓
100m Region 2 [150°E-180°E]x[7°S-7°N]	✓	✓	✓	✓
150m Region 2 [150°E-180°E]x[7°S-7°N]	✓	✓	✓	✓
200m Region 2 [150°E-180°E]x[7°S-7°N]	✓		✓	✓
250m Region 2 [140°E-170°E]x[7°S-7°N]	✓	✓	✓	✓
300m Region 2 [160°E-200°E]x[10°S-3°N]	✓	✓	✓	✓
400m Region 2 [150°E-170°E]x[10°S-0°]	✓	✓	✓	
500m Region 2 [150°E-170°E]x[10°S-0°]	✓	✓	✓	
Zonal wind stress				
WND Region 3 [160°E-200°E]x[0°-10°N]		✓	✓	✓
WND Region 1 [180°E-220°E]x[4°S-4°N]		✓		✓
WND Region 2 [180°E-210°E]x[10°S-0°]		✓		✓

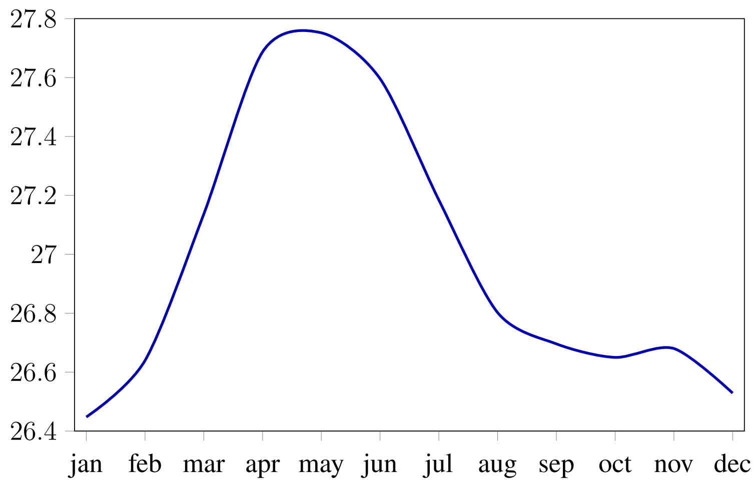

Figure A.1: Average Niño3.4 index per month over the entire sample from January 1982 to December 2024

Table A.4: In-sample parameter estimation results for subsample 2, i.e. 2005–1 to 2024–12, using $q = 2$ and $r = 2$ and for the models with seasonals and cycles. Standard errors are reported in italics. For the models with $r = 2$, $\phi_{\gamma,1}^{(1)}$ and $\phi_{\gamma,2}^{(1)}$ have been fixed at zero.

		Seasonals (q=2)		Seasonals + cycles (q=2,r=2)	
	constant	constant	trend	constant	trend
σ^2	0.967	0.115	0.096	0.085	0.086
	0.089	*0.012*	*0.009*	*0.008*	*0.008*
α	27.015	27.006		27.017	
	0.064	*0.113*		*0.165*	
θ_α			1.127		0.094
			0.076		*0.088*
$\phi_{\gamma,1}^{(1)}$		1.000	0.999	0.997	0.998
		n/a	*n/a*	*n/a*	*n/a*
$\phi_{\gamma,2}^{(1)}$		-1.000	-0.999	-0.993	-0.995
		0.006	*0.006*	*0.010*	*0.009*
$\theta_\gamma^{(1)}$		0.038	0.056	0.086	0.075
		0.016	*0.019*	*0.038*	*0.029*
$\phi_{\gamma,1}^{(2)}$		1.129	1.729	1.732	1.732
		n/a	*n/a*	*n/a*	*n/a*
$\phi_{\gamma,2}^{(2)}$		-0.425	-0.996	-1.000	-1.000
		0.037	*0.006*	*0.004*	*0.004*
$\theta_\gamma^{(2)}$		1.199	0.088	0.057	0.057
		0.071	*0.044*	*0.023*	*0.023*
$\phi_{\psi,1}^{(1)}$				0.000	0.000
				n/a	*n/a*
$\phi_{\psi,2}^{(1)}$				0.000	0.000
				n/a	*n/a*
$\theta_\psi^{(1)}$				0.420	0.370
				0.070	*0.071*
$\phi_{\psi,1}^{(2)}$				1.537	1.545
				0.091	*0.089*
$\phi_{\psi,2}^{(2)}$				-0.629	-0.628
				0.089	*0.089*
$\theta_\psi^{(2)}$				0.644	0.630
				0.097	*0.107*
Periodicity $\gamma_t^{(1)}$		6 (fixed)	6 (fixed)	6 (fixed)	6 (fixed)
Periodicity $\gamma_t^{(2)}$		12 (fixed)	12 (fixed)	12 (fixed)	12 (fixed)
Periodicity $\psi_t^{(1)}$				-	-
Periodicity $\psi_t^{(2)}$				25.124	27.968
LLH	-328.11	-79.28	-58.29	-42.99	-44.84
AIC	660.22	170.56	128.58	105.98	109.68
BIC	667.13	191.30	149.31	140.53	144.23

Table A.5: Selected lags of the explanatory variables for models with untransformed regressors corresponding to results in Table 2 for horizons 1 and 6

	horizon 1			horizon 6		
	constant		seas + cycles	constant		seas + cycles
	LASSO	G2S	G2S	LASSO	G2S	G2S
RB	6, 8, 9			6, 7, 8, 9, 11	6, 11	
WPAC	1, 3, 4				18	
WPAC2	4, 17					
WPAC3		14		16, 17	17	17
WPAC4		14				
50m	5	2, 19, 25			25, 31	
100m "cold"	6, 9, 11, 19, 20, 28, 29	6, 7	25	6, 11, 20, 21, 22, 28, 29, 35, 36	6, 7, 30, 36	25
100m Region 1	1, 32	5, 11, 36	1	6, 19	25	
100m Region 2	1, 5, 6, 8, 10, 28	1, 2, 19, 20	1	7, 8, 10	6, 11, 31, 35	
150m Region 1	1, 2	1, 7, 23, 31, 36		6	6, 19, 31	
150m Region 2	1, 17	2, 5, 13, 21, 25, 26, 36			30	13
200m Region 1	2	10, 11, 29			31	10, 22
200m Region 2	13, 22	2, 19, 34, 36		14	36	
250m Region 1	8, 36	1, 9, 13, 25, 32		36	23	
250m Region 2	12, 13	2, 8, 11, 14		12, 13, 18	13, 18	
300m Region 1		9			6	
300m Region 2	2, 3, 4, 13, 17, 18, 23, 30, 31, 32	1, 2, 13, 17, 25	13, 17	13, 18, 19, 30, 31, 32, 33		
400m Region 1	1, 13, 14, 15	1, 17			13	
400m Region 2	2, 16	1, 10		6, 7		
500m Region 1	3, 4, 8, 25, 26, 32, 35, 36			27, 28, 29, 30, 35	30	
500m Region 2	6, 10, 11	13		10, 11	10, 12	
WND Region 3	22, 23	6, 24, 28		6, 7, 8, 23, 24	28	
WND Region 1	6, 32	24, 28		6, 7, 32	19, 28	
WND Region 2	14, 20, 21, 22, 25, 26, 32, 33	20		14, 25, 26, 31, 32, 33	31	

Table A.6: Selected lags of the explanatory variables for models with untransformed regressors corresponding to results in Table 2 for horizons 12 and 18

	horizon 12			horizon 18		
	constant		seas + cycles	constant		seas + cycles
	LASSO	G2S	G2S	LASSO	G2S	G2S
RB						
WPAC	15				18	
WPAC2	12			18		
WPAC3	14, 15		17			
WPAC4	12					
50m	25	31			20, 25, 31	
100m ”cold”	21, 29, 30, 31, 34, 36	25, 36			36	25
100m Region 1	20	36				
100m Region 2	31		25		20	
150m Region 1	31, 32				25	
150m Region 2	12, 13, 14, 17, 33, 34	25, 36		18	18, 36	
200m Region 1						
200m Region 2	12, 13, 22, 31	15, 25, 36	13		34, 36	
250m Region 1	12, 13, 14, 20, 28, 35, 36	23, 33, 36		24		
250m Region 2	12, 13	13				
300m Region 1	22, 23, 24, 28, 32, 35, 36					
300m Region 2	18, 24, 28, 30, 31, 32, 33, 34, 35	34		18, 33, 34	34	
400m Region 1	12, 14, 17, 18	17		18		
400m Region 2		13				
500m Region 1	27, 29, 30, 31, 32, 33	26, 29		25, 26, 27	25, 29	
500m Region 2	12, 13, 15	13				
WND Region 3	23, 25, 27, 28, 29, 34, 35, 36					
WND Region 1	19, 20, 21, 22, 31, 32	19				
WND Region 2	12, 13, 14, 15, 16, 17, 24, 25, 26, 31, 32, 33, 34, 35, 36	32		32, 33	32	

Table A.7: Selected lags of the explanatory variables for models with PC regressors corresponding to results in Table 2

	horizon 1			horizon 6			horizon 12			horizon 18		
	constant		seas + cycles	constant		seas + cycles	constant		seas + cycles	constant		seas + cycles
	LASSO	G2S	G2S	LASSO	G2S	G2S	LASSO	G2S	G2S	LASSO	G2S	G2S
PC 1	1, 4, 5, 9, 13, 14, 19, 20, 22, 23, 25, 29, 35	1, 5, 6, 12, 13, 25	1, 13, 17	6, 13, 15, 16, 17, 19, 21, 22, 28, 30, 35, 36	19, 36		12, 13, 14, 17, 20, 32, 33	13, 25	17, 24	18, 25, 32, 33	18, 25, 30	
PC 2	1, 2, 4, 8, 9, 10, 13, 18, 23, 25, 26, 31, 32	1, 8, 33		6, 20, 25, 32, 33, 34, 35, 36	6, 25, 35		15, 16, 17, 25, 26, 27, 32, 33, 34, 35, 36	33		25, 26, 32, 33	33	
PC 3	1, 2, 3, 6, 7, 8, 9, 13, 14, 19	2, 5, 6, 19		6, 7, 8, 9, 10, 11, 12, 19, 20, 35, 36	6, 9, 19, 36	14, 36	19, 20	19	14	19		36
PC 4	2, 3, 4, 5, 11, 13, 14, 15, 16, 17, 18, 19, 20, 23, 24, 25	5, 12		7, 8, 9, 10, 11, 13, 14, 16, 19, 31	11, 12	21, 22, 23, 24, 25	12, 13, 24, 32	16, 32				21, 22, 23, 24, 25
PC 5	1, 2, 3, 9, 10, 11, 12, 13, 22, 29, 36	5, 12, 29	14, 29	6, 12, 13, 14, 29	12		12, 13, 14, 15	12, 15				
PC 6	1, 2, 5, 8, 9, 10, 21, 28, 36	6, 8, 35		8, 9, 10, 11, 29	8, 11		12, 23, 26, 29, 32	25			18, 25	

Table A.8: Comparing seasonal-cycles and seasonal-dummies: mean squared forecast errors based on subsample period 2, i.e. 2005–1 to 2024–12. Forecasts in the 90% and 75% model confidence sets are indicated by one and two asterisks, respectively. A constant and low-frequency cycles are included in all specifications; other components are included only when marked by $\times$.

Components				Forecasting horizon h						
Seasonal cycles	Seasonal dummies	G2S Regressors	G2S PCA Regressors	1	3	6	12	14	16	18
$\times$				0.082	0.324**	0.670**	0.820**	0.815**	0.818**	0.819**
	$\times$			0.073**	0.287**	0.623**	0.827**	0.820**	0.818**	0.820**
$\times$	$\times$			0.072**	0.284**	0.615**	0.856**	0.865**	0.864**	0.868**
$\times$		$\times$		0.088	0.347*	0.699**	0.928**	0.862**	0.851**	0.860**
	$\times$	$\times$		0.084**	0.346**	0.694**	0.817**	0.846**	0.882**	0.895**
$\times$	$\times$	$\times$		0.071**	0.285**	0.630**	0.850**	0.866**	0.868**	0.887**
$\times$			$\times$	0.087	0.328**	0.674**	0.844**	0.860**	0.864**	0.847**
	$\times$		$\times$	0.076**	0.318**	0.660**	0.932**	0.935**	0.881**	0.845**
$\times$	$\times$		$\times$	0.072**	0.285**	0.634**	0.856**	0.868**	0.868**	0.869**

Table A.9: Mean squared forecast errors for El Niño episodes (Jan 2005–Mar 2005, Aug 2006–Feb 2007, Jun 2009–Apr 2010, Sep 2014–May 2016, Aug 2018–Jul 2019, Apr 2023–May 2024). Forecasts in the 90% and 75% model confidence sets are indicated by one and two asterisks, respectively.

	Forecasting horizon h						
	1	3	6	12	14	16	18
Constant	1.474	1.546	1.599	1.636**	1.636**	1.630**	1.621**
Regressors							
G2S	0.328	0.902	1.633**	2.047*	1.799**	1.915*	2.252
G2S PCA	0.497	0.799*	1.403**	1.970*	1.781**	1.624**	1.892
LASSO	0.296	0.867	1.243**	1.724**	1.745**	1.909	1.874
LASSO PCA	0.335	0.884	1.394*	1.822**	1.784**	1.737**	1.850
Seasonals							
constant	0.112	0.816	1.608*	1.584**	1.590**	1.590**	1.591**
trend	0.104	0.554**	1.284**	2.243*	2.419	2.576	2.697
Seasonals + cycles							
constant	0.096	0.524*	1.273**	1.776**	1.774**	1.757*	1.731
trend	0.094	0.491**	1.211**	1.877*	1.930	1.965	1.948
Seasonals + cycles + regressors							
G2S	0.099*	0.555*	1.301**	1.915*	1.894	1.834*	1.820**
G2S PCA	0.104	0.521**	1.287**	1.859*	1.892	1.869*	1.803
Anomalies							
AR(1)	0.102	0.482**	1.119**	1.890*	1.907	1.923	1.938
AR(3)	0.079**	0.420**	1.122**	1.747**	1.741**	1.735*	1.730
Anomalies + regressors							
ARX(1) G2S	0.083**	0.435**	1.288**	1.491**	1.538**	1.395**	1.331**
ARX(1) G2S PCA	0.102	0.544**	1.658	1.806**	1.641**	1.772*	1.773
ARX(3) G2S	0.076**	0.551*	1.348**	1.551**	1.554**	1.552**	1.653**
ARX(3) G2S PCA	0.091	0.427**	1.126**	1.687**	1.633**	1.654**	1.612**

Table A.10: Mean squared forecast errors for La Niña episodes (Oct 2005–Apr 2006, May 2007–Jul 2008, Oct 2008–Apr 2009, May 2010–May 2012, Jul 2016–Jan 2017, Sep 2017–May 2018, and Jul 2020–Feb 2023). Forecasts in the 90% and 75% model confidence sets are indicated by one and two asterisks, respectively.

	Forecasting horizon h						
	1	3	6	12	14	16	18
Constant	0.993	1.010	1.017	1.026	1.020	1.029	1.028
Regressors							
G2S	0.230	0.707	0.908*	1.612	1.471	1.177	0.970*
G2S PCA	0.219	0.478	0.578**	0.933	0.798*	0.683**	0.640**
LASSO	0.160	0.392	0.644**	1.001	0.767**	0.663**	0.690**
LASSO PCA	0.170	0.487	0.680**	0.782	0.724**	0.662**	0.713*
Seasonals							
constant	0.109	0.597	0.967	0.989	0.988	1.017	1.024
trend	0.080	0.442	1.391	1.730	1.473	1.316	1.405*
Seasonals + cycles							
constant	0.065*	0.247	0.580**	0.632**	0.618**	0.631**	0.639**
trend	0.067	0.259	0.618**	0.795	0.761*	0.729**	0.722*
Seasonals + cycles + regressors							
G2S	0.071	0.264	0.606**	0.717	0.631**	0.635**	0.643**
G2S PCA	0.070	0.255	0.577**	0.633**	0.646**	0.664**	0.645**
Anomalies							
AR(1)	0.068*	0.291	0.813**	1.034	0.892*	0.767**	0.742*
AR(3)	0.057**	0.204**	0.559**	0.673**	0.643**	0.625**	0.622**
Anomalies + regressors							
ARX(1) G2S	0.070	0.306	0.967*	1.240	0.999*	0.881**	0.886*
ARX(1) G2S PCA	0.076	0.329	1.022	1.611	1.433	0.952*	0.923*
ARX(3) G2S	0.073	0.322	0.774*	0.977	0.906*	0.752**	0.739**
ARX(3) G2S PCA	0.067	0.247	0.615**	0.843	0.716**	0.607**	0.633**

Table A.11: Mean squared forecast errors based on subsample period 2, i.e. 2005–1 to 2024–12, without anomaly benchmarks. Forecasts in the 90% and 75% model confidence sets are indicated by one and two asterisks, respectively.

	Forecasting horizon h						
	1	3	6	12	14	16	18
Constant	0.971	0.997	1.010	1.035	1.035	1.040	1.037
Regressors							
G2S	0.292	0.778	1.117	1.544	1.520	1.328	1.284
G2S PCA	0.317	0.507	0.748**	1.051*	0.967*	0.881**	0.932**
LASSO	0.205	0.491	0.708**	1.029*	0.923**	0.936*	0.936**
LASSO PCA	0.228	0.521	0.760**	0.946*	0.913**	0.872**	0.925**
Seasonals							
constant	0.112	0.577	0.967	0.983*	0.987*	1.004*	1.010
trend	0.092	0.457	1.177	1.615	1.527	1.563	1.688
Seasonals + cycles							
constant	0.082**	0.324**	0.670**	0.820**	0.815**	0.818**	0.819**
trend	0.083**	0.327**	0.680**	0.928*	0.923*	0.920*	0.910**
Seasonals + cycles + regressors							
G2S	0.088*	0.347**	0.699**	0.928*	0.862**	0.851**	0.860**
G2S PCA	0.087*	0.328**	0.674**	0.844**	0.860**	0.864**	0.847**

Table A.12: Out-of-sample forecasts of the Niño3.4 index made on October 2024 corresponding to the plots in Figure 6.

	$h = 1$ 11-'24	$h = 3$ 1–'25	$h = 6$ 4–'25	$h = 12$ 10-'25	$h = 14$ 12-'25	$h = 16$ 2–'26	$h = 18$ 4–'26
Levels							
Only constant	27.0290	27.0290	27.0290	27.0290	27.0290	27.0290	27.0290
Regressors							
G2S	26.4491	26.1909	27.4686	25.1546	24.7056	24.0669	25.8105
G2S PCA	26.7499	26.0044	27.1743	25.8109	25.1194	25.6022	27.2768
LASSO	26.2116	25.9475	27.3638	26.1915	24.9219	25.7679	27.0417
LASSO PCA	26.4314	25.9696	26.9573	26.7234	26.2583	25.5713	26.8765
Seasonals							
constant	26.4554	26.5552	27.2086	27.1443	27.1070	26.7630	27.1493
trend	26.3754	26.3451	27.3957	26.3724	26.3398	26.5592	27.3872
Seasonals + cycles							
constant	26.3637	26.3868	27.5588	26.6110	26.5527	26.8056	27.6378
trend	26.3739	26.4183	27.6387	26.7401	26.6881	26.9455	27.7811
Seasonals + cycles + regressors							
G2S	26.2877	26.3325	27.5588	26.6890	26.5716	26.7278	27.6297
G2S PCA	26.3637	26.3868	27.5902	26.6110	26.5527	26.8056	27.6378
Anomalies							
Only constant	0.3488	0.5814	-0.6580	0.3790	0.5002	0.3895	-0.6580
Regressors							
G2S	-0.2311	-0.2567	-0.2184	-1.4954	-1.8232	-2.5726	-1.8766
G2S PCA	0.0697	-0.4432	-0.5128	-0.8391	-1.4094	-1.0374	-0.4102
LASSO	-0.4686	-0.5001	-0.3232	-0.4586	-1.6069	-0.8716	-0.6453
LASSO PCA	-0.2488	-0.4780	-0.7297	0.0734	-0.2705	-1.0682	-0.8105
Seasonals							
constant	-0.2248	0.1076	-0.4784	0.4942	0.5782	0.1235	-0.5377
trend	-0.3048	-0.1025	-0.2913	-0.2776	-0.1890	-0.0804	-0.2998
Seasonals + cycles							
constant	-0.3165	-0.0608	-0.1282	-0.0390	0.0239	0.1661	-0.0492
trend	-0.3063	-0.0293	-0.0483	0.0900	0.1593	0.3059	0.0941
Seasonals + cycles + regressors							
G2S	-0.3925	-0.1151	-0.1282	0.0390	0.0428	0.0883	-0.0573
G2S PCA	-0.3165	-0.0608	-0.0969	-0.0390	0.0239	0.1661	-0.0492

Note: The bottom panel reports forecasted anomalies obtained by subtracting the monthly sample mean of the corresponding month.

B Details of statistical model, estimation and forecasting

The univariate time series model for the Niño3.4 index in month t is given by

$$y_t = \mu_t + x_t'\beta + \varepsilon_t, \qquad \varepsilon_t \sim IID(0, \sigma^2), \qquad t = 1, \ldots, T,$$

where μ_t is the signal variable, x_t is the $k \times 1$ vector of lagged covariates, fixed and known, in month t, β is the corresponding $k \times 1$ vector of fixed regression coefficients, and ε_t is the noise variable with mean zero and variance $\sigma^2 > 0$. We notice that T is the length of the sample and k is the dimension of the vector of covariates x_t.

B.1 Signal specifications

In case μ_t is constant over time, that is $\mu_t = \alpha$ for all $t = 1, \ldots, T$, and α is assumed to be a fixed unknown coefficient, model equation (1) reduces to the standard linear regression model. In case we treat μ_t as a time-varying signal variable, we formulate μ_t as a dynamic statistical model. In our study, we treat μ_t as a variable that is stochastically evolving over time and we specify it accordingly as a dynamic equation. To have some level of flexibility in the model specification, while still remaining in a linear framework, we propose to decompose μ_t into specific dynamic variables that can account for the long-term dynamics (trend), the within-year periodic effects (seasonal), and the stationary dynamics (cycles). In particular, we specify the signal as

$$\mu_t \quad = \quad \alpha_t + \gamma_t + \psi_t,$$

where α_t represents the constant or trend, γ_t is the seasonal effect, and ψ_t is the cycle variable. This model resembles the score-driven component models of Harvey and Luati (2014) and Caivano et al. (2016), although these only feature a trend and a seasonal component. The trend component is formulated as a non-stationary random walk process as given by

$$\alpha_{t+1} = \alpha_t + \theta_\alpha s_t,$$

where θ_α is a fixed scale parameter, which can be interpreted as the signal-to-noise coefficient for the trend, and where s_t is a zero-mean noise term that is defined below. We notice that in the case

$\theta_\alpha = 0$, the trend component α_t reduces to the constant $\alpha = \alpha_1$ as defined above.

The seasonal γ_t and cycle ψ_t components are modeled as sums of stationary autoregressive (AR) processes with a maximum lag of two time periods. We have

$$\gamma_t = \gamma_t^{(1)} + \ldots + \gamma_t^{(s)}, \qquad \psi_t = \psi_t^{(1)} + \ldots + \psi_t^{(r)},$$

where each sub-component is modeled as an AR(2) process, that is

$$\gamma_{t+1}^{(i)} = \phi_{\gamma,1}^{(i)} \gamma_t^{(i)} + \phi_{\gamma,2}^{(i)} \gamma_{t-1}^{(i)} + \theta_\gamma^{(i)} s_t, \qquad i = 1, \ldots, s,$$

and

$$\psi_{t+1}^{(j)} = \phi_{\psi,1}^{(j)} \psi_t^{(j)} + \phi_{\psi,2}^{(j)} \psi_{t-1}^{(j)} + \theta_\psi^{(j)} s_t, \qquad j = 1, \ldots, r,$$

with autoregressive coefficients $\phi_{\gamma,p}^{(i)}$ and $\phi_{\psi,p}^{(j)}$, for $p = 1, 2$, and with scale coefficients $\theta_\gamma^{(i)}$ and $\theta_\psi^{(j)}$, which are treated as fixed parameters. We notice that all AR(2) processes share the same noise term s_t, which is a characteristic inherent to score-driven models, as discussed in more detail below. The relative importance of each sub-component in the dynamic specification of the signal μ_t is mostly determined by the scale coefficients θ, which are assumed to be strictly positive. For all AR(2) processes, we set the initial values of $\gamma_1^{(i)}$ and $\psi_1^{(j)}$ to zero, which are their unconditional expectations, given that they belong to stationary processes. Whenever the filters for these components are invertible, see e.g. Blasques et al. (2018), the effect of the starting values vanishes quickly so they cannot be consistently estimated. The numbers of sub-components for the seasonal (q) and the cycle (r) components can be determined using different methodologies, including those based on information criteria (such as Akaike information criterion, AIC) and on minimizing particular loss functions computed using sub-samples from the data. More details on determining q and r are provided below.

To address seasonal fluctuations in the Niño3.4 time series, we base the seasonal component γ_t on the sum of AR(2) processes. The specification does not involve the inclusion of seasonal lags (that is, in the case of the monthly Niño3.4 index time series, multiples of 12 lags). Instead, seasonal sub-components are formulated as AR(2) processes with AR coefficients $\phi_{\gamma,1}^{(i)}$ and $\phi_{\gamma,1}^{(i)}$, for $i = 1, \ldots, s$, being constrained such that the roots of the corresponding characteristic equations, $1 - \phi_{\gamma,1}^{(i)} z - \phi_{\gamma,1}^{(i)} z^2 = 0$, are in the complex range and the implied cyclical processes $\gamma_t^{(i)}$ have a frequency of around $2\,\pi\,/\,6\,i$. Such restrictions impose periodicities within the cyclical processes of 6 and 12 months for $i = 1$ and $i = 2$, respectively; see Appendix B.2 for details of how these restrictions are imposed. This specification of the seasonal component γ_t can be extended by including more AR lags, but we will not explore this further. In case of the Niño3.4 time series, empirical evidence suggests that its seasonal fluctuation can be captured well by these two seasonal sub-components, that is $q = 2$, see the discussion in Petrova et al. (2020).

The cycle component ψ_t and its sub-components are handled in a similar way to the seasonal component, that is, by imposing complex roots for the AR polynomials, but without imposing the periodicity within a year span. The periodicity of the cycle is estimated for each sub-component $\psi_t^{(j)}$ of the cycle ψ_t. In practice, and to facilitate the interpretation of the sub-cycles, an admissible range for the periodic length is provided. In this way, a cycle sub-component can be associated with a particular short-, medium-, or long-term dynamic feature of ENSO.

B.2 Imposing cyclicality of autoregressive processes

This section describes how we enforce cyclical dynamics on the seasonal and cyclical components $\gamma_t^{(i)}$ and $\psi_t^{(j)}$, and how we restrict their periodicity. These components are AR(2) processes, so let us consider a general AR(2) model:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t\,.$$

Processes generated from this model exhibit cyclical behavior if the roots of the lag polynomial $A(z) = 1 - \phi_1 z - \phi_2 z^2$ are complex. Instead of solving for the roots of $A(z)$, it is simpler to find the roots of the reciprocal polynomial $z^2 - \phi_1 z - \phi_2$, which correspond to the inverse roots of $A(z)$. These roots are given by:

$$z_1, z_2 = \frac{1}{2}\left(\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}\right)\,.$$

So, to enforce complexity of the roots, we must have:

$$\phi_1^2 + 4\phi_2 < 0\,,$$

which requires $\phi_2 < 0$. In case this case, the moduli of the inverse roots of the characteristic polynomial are equal to $\sqrt{-\phi_2}$, so $0 > \phi_2 > -1$ is needed to have stationarity and cyclicality of the process. Stationarity of the seasonal components is discussed in Section 2.1. Furthermore, it is a standard result that the frequency of the cycle is given by:

$$\frac{1}{2\pi}\cos^{-1}\left(\frac{\phi_1}{2\sqrt{-\phi_2}}\right).$$

The periodicity τ is the inverse of the frequency, so ϕ_1 can be expressed as a function of τ and ϕ_2:

$$\phi_1 = 2\sqrt{-\phi_2}\cos(2\pi/\tau).$$

Thus, instead of optimizing ϕ_1 and ϕ_2 directly, we can reparameterize the cyclical AR(2) model by optimizing over τ and ϕ_2, reconstructing ϕ_1 as above. By enforcing $0 > \phi_2 > -1$, we ensure stationarity and cyclicality, while restricting τ allows control over the periodicity of the cycle.

B.3 Score-driven model and filter

All dynamic sub-components are driven by the same zero-mean noise sequence s_t, which defines the model (1) as a score-driven model, see Creal et al. (2013) and Harvey (2013). In particular, we define s_t as the score function of the predictive density contribution at time t, as denoted by $p(y_t|y_1,\ldots,y_{t-1},x_1,\ldots,x_t;\lambda)$, where λ is the parameter vector containing all unknown fixed coefficients in model (1), including β, with respect to the time-varying signal μ_t. More specifically, we define s_t as

$$s_t = S_t\,\frac{\partial\log p(y_t|y_1,\ldots,y_{t-1}x_1,\ldots,x_t;\lambda)}{\partial\mu_t},$$

where S_t is a scaling function such as the inverse of the corresponding information or a function thereof. For all practical purposes, one can also set S_t equal to unity, $S_t = 1$. A convenient feature of the score-driven model is that the score sequence $\{s_t\}$ constitutes a martingale difference. The functional form of the score function depends on the assumed density function of the disturbance ε_t in the model equation (1), since $p(y_t|y_1,\ldots,y_{t-1}x_1,\ldots,x_t;\lambda) = p_\varepsilon(\varepsilon_t;\lambda)$, where $\varepsilon_t = y_t - \mu_t - x_t'\beta$. In case of the normal density, the score function is simply given by $s_t = \varepsilon_t$, with the scale $S_t = \sigma^2$ set equal to the inverse of the conditional Fisher information of μ_t. When we apply this score-driven framework to the Niño3.4 time series, for a given set of covariates, and for given values of the parameters in λ, the scores s_t become actual values and the time-varying (sub-)components can be constructed recursively according to the equations given above. Hence, in this case, the formulation of the score-driven model becomes a set of filtering equations.

Using the score to update the time-varying (sub-)components is intuitive, because the score is the steepest ascent direction to improve the local fit in terms of log likelihood at time t, given the current position of the time-varying parameter μ_t. Thus, the score provides a natural direction for updating the (sub-)components. We notice that this holds regardless of whether the model is correctly specified or not. The conditions for optimality of the score-driven filter are documented in Blasques et al. (2015). When the score framework is applied to particular densities and particular parameters, many well-known dynamic models are obtained as a result; see Artemova et al. (2022).

B.4 Parameter estimation and forecasting

The econometric methodology for score-driven models is set out in various papers; an overview is given by Artemova et al. (2022). Next, we provide and discuss the key details that are relevant for our current model and implementation. We estimate the parameters using the method of maximum likelihood. The log likelihood depends on the values of the components α_t, γ_t, and ψ_t. As they are not observed, we have to filter them using their updating equations. We denote a filtered version of one of these processes by putting a hat on top of it, e.g. $\widehat{\alpha}_t(\lambda)$ denotes the filtered version of α_t for a given parameter λ. This value should not be confused with the 'true' value of α_t, which depends on some true λ_0. The maximum likelihood problem is then given by

$$\hat{\lambda}_T = \arg\max_{\lambda} \sum_{t=1}^{T} \ell_t(\lambda) \, ,$$

where

$$\begin{aligned}
\ell_t(\lambda) &= \log p(y_t | y_1, \ldots, y_{t-1}, x_1, \ldots, x_t; \lambda) \\
&= \log p_\varepsilon(y_t - \widehat{\mu}_t(\lambda) - x_t'\beta; \lambda) \\
&= \log p_\varepsilon(y_t - \widehat{\alpha}_t(\lambda) - \widehat{\gamma}_t^{(1)}(\lambda) - \ldots - \widehat{\gamma}_t^{(s)}(\lambda) - \widehat{\psi}_t^{(1)}(\lambda) - \ldots - \widehat{\psi}_t^{(r)}(\lambda) - x_t'\beta; \lambda) \, .
\end{aligned}$$

The score-driven filters for the stationary components $\gamma_t^{(i)}$ and $\psi_t^{(i)}$ are initialized at zero, which is their unconditional mean. The stochastic trend term α_t is initialized at the average value of the first six observations. When we exclude a stochastic trend setting $\theta_\alpha = 0$, we include the constant value α in the parameter vector λ and estimate it along with the other parameters. To allow the score-driven filters to converge to the correct trajectories, we use a burn-in period of 6 months, which means that we exclude the first 6 log likelihood terms from the optimization. To ensure identification of the different components and to improve stability of the optimization routine, we restrict the cycle frequency (associated with the cycle periodicity) to lie in certain intervals; see also the discussion above. We enforce non-explosive behavior of the seasonal and cycle processes by appropriately transforming the coefficients in such a way that the roots of the characteristic polyomial of the AR processes are never strictly inside the unit circle.

The conditions for consistency and asymptotic normality of the maximum likelihood estimator under stationarity are derived in Blasques et al. (2022). Since under Gaussianity our model is linear, these conditions can be verified straightforwardly. When the model includes the stochastic trend α_t, it is unit root non-stationary. In Blasques et al. (2024), it is shown that even under unit root non-stationarity, location models of the type we consider here remain to have regular asymptotic properties. The reason is that the log likelihood contributions are functions of only the prediction errors and static parameters. These errors are asymptotically stationary, even under model misspecification, because the unit root is canceled out in the computation of the prediction error $y_t - \alpha_t - \gamma_t - \psi_t$. For further details, we refer to Blasques et al. (2024).

Finally, we briefly describe how the score-driven model is used for forecasting. Given that s_T is known at time T, $\hat{\mu}_{T+1}(\widehat{\lambda}_T)$ can be evaluated based on the filtering equations. Thus, one step ahead forecasting is trivial. For larger forecast horizons, we use $\mathbb{E}_T[s_{T+k}] = 0$ for $k > 1$, where $\mathbb{E}_T[\cdot]$ denotes the conditional expectation given the information available at time T, since $\{s_t\}$ is a martingale difference sequence. Thus, forecasts for more than one step ahead can be obtained by recursively evaluating the filtering equations with 0 filled in place of s_{T+k}. For regressors, we only consider lag orders that are equal to or greater than the forecast horizon h. The term $x'_{T+h}\beta$ is therefore known at time T and can be used directly for forecasting.