Computational Statistics and Data Analysis 56 (2012) 3134–3152

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling

Virginie Dordonnat [a], Siem Jan Koopman [b,*], Marius Ooms [b]

[a] Électricité de France, Research & Development, Clamart, France

[b] VU University Amsterdam, De Boelelaan 1105, NL-1081 HV Amsterdam, Netherlands

ARTICLE INFO

Article history:
Available online 9 April 2011

Keywords:
Kalman filter
Electricity load forecasting
Forecasting
Periodic regression
Signal extraction
Time-varying parameter

ABSTRACT

A dynamic multivariate periodic regression model for hourly data is considered. The dependent hourly univariate time series is represented as a daily multivariate time series model with 24 regression equations. The regression coefficients differ across equations (or hours) and vary stochastically over days. Since an unrestricted model contains many unknown parameters, an effective methodology is developed within the state–space framework that imposes common dynamic factors for the parameters that drive the dynamics across different equations. The factor model approach leads to more precise estimates of the coefficients. A simulation study for a basic version of the model illustrates the increased precision against a set of univariate benchmark models. The empirical study is for a long time series of French national hourly electricity loads with weather variables and calendar variables as regressors. The empirical results are discussed from both a signal extraction and a forecasting standpoint.

1. Introduction

This paper develops a general method to analyse common dynamic features in multivariate time-varying regression models. We further investigate parameter changes for high frequency periodic time series models. The main idea is to introduce dynamic factor models for individual time series with similar characteristics. In our application, we analyse a daily vector time series; each time series is associated with an hour of the day. The aim is then to find common dynamic features in the time-varying regression coefficients for different hours. We discuss the implementation of a dynamic factor state–space model, including time-varying coefficients for trend, seasonal and regression effects. We show that the method can be successful for a long time series of hourly electricity loads where we take account of stochastic trends, yearly cycles, calendar effects and the changing influence of temperature. To emphasise its practical relevance, we compare our dynamic factor model with univariate benchmark models, both for signal extraction and forecasting. We concentrate on the forecasting of hourly electricity loads for regular days, including weekend days.

The challenge of modelling high frequency periodic time series is the detection of the recurring but persistently changing patterns within days, weeks and years. Some patterns are more variable than others and imply different forecasting functions. Fixed patterns can be exploited for long forecast horizons, whereas variable patterns are more relevant for short-term forecasts. Time-varying regression models provide a convenient statistical framework to tackle the problem. Regression models in the context of time series may contain a constant, seasonal effects and other explanatory variables. When the associated regression coefficients are allowed to change over time, the result is a flexible methodology that transforms the mean into a trend function and enables tracking time-varying patterns in the seasonal effects.

* Corresponding author.
E-mail address: s.j.koopman@vu.nl (S.J. Koopman).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2011.04.002

In case of high-frequency seasonal time series, the challenge of signal extraction and forecasting is even higher. The confounding of different seasonal effects in the same time series become apparent since, for example, daily time series are subject to quarterly effects (summer, winter), day-of-the-week effects (weekday, weekend), calendar effects (Christmas, Easter) and, possibly, weather effects. Such issues become particularly important when the time series to forecast corresponds to hourly measurements. An example in the context of electricity loads is given by Harvey and Koopman (1993) where hourly loads are forecast using time-varying regression smoothing splines. Given the typical noisy structure of the time series, the large number of recurring effects that need to be considered in the analysis and the vast amount of available data, forecasting electricity loads is widely seen as a challenging task.

A review of load forecasting methods is given by Bunn and Farmer (1985) and the references therein. More up-to-date reviews of the literature are provided by Lotufo and Minussi (1999), Weron (2006) and Hahn et al. (2009). Taylor and McSharry (2007) discuss short-term load forecasting (one-hour to one-day ahead) using standard forecasting methods including seasonal autoregressions and exponential smoothing which are also carried out in a periodic fashion by treating each hourly load as a daily time series. Advanced methods for load forecasting are developed by Cottet and Smith (2003) who adopt Bayesian procedures for forecasting high-dimensional vectors of time series; see also Panagiotelis and Smith (2010) and Ohtsuka et al. (2010) for more recent illustrations. The covariance structures in such multivariate time series are of key importance for an effective forecasting strategy. Both Smith and Kohn (2002) and Cottet and Smith (2003) take account of the correlation between hourly loads when computing their forecasts. Espinoza et al. (2005) analyse many time series at different grid points separately using periodic models in order to find commonalities in the identified characteristics of the time series. In this way, common profiles in time series are formulated which form the basis for the joint forecasting of the time series. The internal statistical model of Electricité de France is described in Bruhns et al. (2005) and is primarily developed for forecasting the French hourly load. Our modelling framework is useful for signal extraction of changes in regression effects and for forecasting. Our approach is based on dynamic factor models since we expect that the dynamic relation between electricity demand and, say, temperature is similar at the hours of, say, 3, 4 and 5 AM. Imposing a common factor for the corresponding parameters may lead to a more parsimonious model; it may also provide a more robust forecast function since information is shared amongst a set of related hours.

In our approach we adopt the econometric approach of Engle and Watson (1981), who specify a one-factor model with constant factor loadings for a multivariate time series. The resulting analysis is based on state–space modelling of multivariate unobserved component models as introduced in Chapter 8 of Harvey (1989). For the estimation of the model parameters, we use a combination of two methods. Whereas Watson and Engle (1983) used a scoring algorithm and an EM (Expectation Maximisation) algorithm, we combine the EM method of Shumway and Stoffer (1982) with quasi-Newton methods for likelihood Maximisation; see also Koopman and Shephard (1992).

Dynamic factor models are also widely used for multivariate macroeconomic data. We will not review all contributions in this literature. A flavor of this work is given by Giannone et al. (2006) who compare vector autoregressive and dynamic factor models for the estimation of business cycles and Del Negro and Otrok (2008) who analyse business cycles based on a factor model with time-varying factor loadings. We find little evidence of macroeconomic effects on daily electricity consumption as our hourly data set is not long enough to cover a full business cycle. Finally we like to mention that the state–space methodology for nonstationary dynamic factor models is also adopted outside economics. For example, Ortega and Poncela (2005) propose a dynamic factor model for fertility rates while Muñoz Carpena et al. (2005) use a dynamic factor model for groundwater quality trends.

The remainder of the paper is organised as follows. Section 2 formulates our model and discusses the state–space framework and implementation details. Section 3 presents a detailed simulation study to illustrate the advantages of dynamic factor modelling for time-varying-parameter estimation and therefore for signal extraction of the temperature effect on electricity loads. Our application to French national hourly electricity loads is presented in Section 4. Section 5 concludes.

2. Dynamic factor regression models for periodic time series

We develop a model for univariate time series subject to seasonal fluctuations associated with different seasonal periods and subject to additional periodic time-varying regression effects. We first concentrate on the shortest seasonal period S. We are specifically interested in the case of high frequency data, where S is relatively large: e.g. in the case of hourly data $S = 24$, in the case of half-hourly data, $S = 48$. Following the periodic time series literature as in Tiao and Grupe (1980) and Hindrayanto et al. (2010), we first transform the univariate time series to a $S \times 1$ vector time series $y_t = \left(y_{1,t} \dots y_{S,t}\right)'$, $t = 1, \dots, T$. In the case of hourly data, the time series regression model for the daily vector y_t implies a periodic model for the hourly series. The general time-varying regression model for y_t we consider is written as:

$$y_t = \mu_t + \sum_{k=1}^{K} B_t^k x_t^k + \varepsilon_t, \quad \varepsilon_t \sim IIN\left(0, \Sigma_\varepsilon\right), \ t = 1, \dots, T, \tag{1}$$

where $\mu_t = \left(\mu_{1,t} \dots \mu_{S,t}\right)'$ is the $S \times 1$ vector of trend components, which captures the smooth long-term evolution of y_t. The disturbance vector ε_t is independently, identically and normally (IIN) distributed with variance matrix Σ_ε. The observations y_t also depend on explanatory variables $x_t^k = (x_{1,t}^k \dots x_{S,t}^k)'$, $k = 1, \dots, K$, which are vector transformations of the original