International Journal of Forecasting 29 (2013) 88–99

Contents lists available at SciVerse ScienceDirect

International Journal of Forecasting

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

Does the Box–Cox transformation help in forecasting macroeconomic time series?

Tommaso Proietti [a,b,*], Helmut Lütkepohl [c,d,1]

[a] University of Sydney, Australia
[b] Università di Roma "Tor Vergata", Italy
[c] Free University Berlin, Germany
[d] DIW Berlin, Germany

ARTICLE INFO

Keywords:
Forecast comparisons
Multi-step forecasting
Rolling forecasts
Nonparametric estimation of prediction error variance

ABSTRACT

The paper investigates whether transforming a time series leads to an improvement in forecasting accuracy. The class of transformations that is considered is the Box–Cox power transformation, which applies to series measured on a ratio scale. We propose a nonparametric approach for estimating the optimal transformation parameter based on the frequency domain estimation of the prediction error variance, and also conduct an extensive recursive forecast experiment on a large set of seasonal monthly macroeconomic time series related to industrial production and retail turnover. In about a fifth of the series considered, the Box–Cox transformation produces forecasts which are significantly better than the untransformed data at the one-step-ahead horizon; in most cases, the logarithmic transformation is the relevant one. As the forecast horizon increases, the evidence in favour of a transformation becomes less strong. Typically, the naïve predictor that just reverses the transformation leads to a lower mean square error than the optimal predictor at short forecast lead times. We also discuss whether the preliminary in-sample frequency domain assessment conducted here provides reliable guidance as to which series should be transformed in order to improve the predictive performance significantly.

1. Introduction

Transformations aim to improve the statistical analysis of time series, by finding a suitable scale for which a model belonging to a simple and well known class, e.g. the normal regression model, has the best performance. An important class of transformations which are suitable for time series measured on a ratio scale with strictly positive support is the power transformation; originally proposed by Tukey (1957) as a device for achieving a model with a simple structure, normal errors and a constant error variance, it was subsequently modified by Box and Cox (1964).

The objective of this paper is to assess whether transforming a variable leads to an improvement in the forecasting accuracy. This issue has already been debated in the time series literature. The use of the Box–Cox transformation as a preliminary specification step before fitting an ARIMA model was recommended in the book by Box and Jenkins (1970). In his discussion of the paper by Chatfield and Prothero (1973), Tunnicliffe Wilson (1973) advocated its use and showed that for the particular case study considered in the paper, the monthly sales of an engineering company, maximum likelihood estimation of the power transformation parameter could lead to superior forecasts. This point was elaborated further by Box and Jenkins (1973).

* Correspondence to: Merewether Building (H04), Discipline of Business Analytics, The University of Sydney, NSW 2006, Australia.
E-mail addresses: t.proietti@econ.usyd.edu.au (T. Proietti), hluetkepohl@diw.de (H. Lütkepohl).
1 DIW Berlin, Mohrenstr. 58, 10117 Berlin, Germany.

0169-2070/$ – see front matter © 2012 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2012.06.001

A more extensive investigation was carried out by Nelson and Granger (1979), who considered a dataset consisting of twenty-one time series. After fitting a linear ARIMA model to the power transformed series and using 20 observations for a post-sample evaluation, they concluded that the Box–Cox transformation does not lead to an improvement in the forecasting performance. Another important conclusion, which is also supported by simulation evidence, is that the naïve forecasts, which are obtained by simply reversing the power transformation, perform better than the optimal forecasts based on the conditional expectation. The explanation for this is that the conditional expectation underlying the optimal forecast assumes that the transformed series is normally distributed. However, this assumption may not be realistic. In contrast to Nelson and Granger's results, Hopwood, McKeown, and Newbold (1981) find, for a range of quarterly earnings-per-share series, that the Box–Cox transformation can improve forecast efficiency.

In related work, Lütkepohl and Xu (2011) have investigated whether the logarithmic transformation (as a special case of a power transformation) leads to an improved forecasting accuracy relative to the untransformed series; the target variables are annual inflation rates, computed from seasonally unadjusted price series. The overall conclusion is that forecasts based on the original variables are characterized by a lower mean square forecast error. On the other hand, based on data on a range of monthly stock price indices, as well as quarterly consumption series, Lütkepohl and Xu (2012) conclude that using logs can be quite beneficial for forecasting. They also point out that there does not appear to be any reliable criterion for deciding between logs and levels for the purpose of maximizing the forecast accuracy.

From a theoretical standpoint, Granger and Newbold (1976) provided a general analytical approach to forecasting transformed series, based on the Hermite polynomials series expansion. Analytical expressions for the minimum mean squared error predictors were provided by Pankratz and Dudley (1987) for specific values of the Box–Cox power transformation parameter. Guerrero (1993) suggested a simple approximate method for obtaining bias-corrected forecasts in the original measurement scale. Carroll and Ruppert (1981) dealt with the contribution of transformation parameter estimation to the overall mean square forecast error, concluding that it is usually small. Collins (1991) discussed and compared different methods for the interval forecasting of transformed series. More recently, Pascual, Romo, and Ruiz (2005) proposed a bootstrap procedure for constructing prediction intervals for a series when an ARIMA model is fitted to its power transformation. De Bruin and Franses (1999) considered forecasting power transformed series using a class of nonlinear time series models (smooth transition autoregressive models). Finally, the Box–Cox transformation is popular in financial time series analysis and has been considered for forecasting volatility (see e.g. Goncalves & Meddahi, 2011; Higgins & Bera, 1992; and Sadorsky & McKenzie, 2008) and price durations (Fernandes & Grammig, 2006), for example.

This paper contributes to the debate in two ways: first, we propose a fast nonparametric method, based on the estimation of the prediction error variance (p.e.v.) of the normalized Box–Cox power transformation, which can be used to estimate the transformation parameter and to decide whether or not to use the power transformation if forecasting is the objective. Our procedure has the advantage that it does not require the normality assumptions which are used in maximum likelihood procedures. Hence, it circumvents the problem observed by Nelson and Granger (1979). Our second contribution is to assess the empirical relevance of the choice of the transformation parameter by performing a large scale recursive forecast exercise, on a dataset consisting of 530 seasonal monthly time series. In the previous studies, only much more limited datasets were used, and by considering such a large dataset we hope to get a better overall picture of the situation, and may be able to explain some of the previous discrepancies in results. A side issue is whether the naïve predictor outperforms the optimal predictor in terms of the mean square forecast error. We find that there is a certain percentage of series where significant forecast improvements are obtained by a power transformation. The challenge is then to identify the series for which a power transformation may help.

The plan of the paper is as follows. In Section 2, after reviewing the Box–Cox transformation, we discuss the predictors of interest. In Section 3, we present the nonparametric procedure for estimating the p.e.v. and the transformation parameter. Section 4 provides a detailed discussion of the advantages and limitations of the method, in the light of the assumptions underlying the analysis. Section 5 discusses the estimation results on the dataset. In Section 6 we judge the relevance of the transformation for out-of-sample forecasting by conducting a rolling forecasting experiment; and conclusions are drawn in Section 7.

2. Forecasting Box–Cox transformed series

Box and Cox (1964) proposed a transformation of a time series variable $y_t,\ t = 1, \dots, n$, that depends on the power parameter λ in the following way:

$$y_t(\lambda) = \begin{cases} \dfrac{y_t^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \ln y_t, & \lambda = 0, \end{cases} \tag{1}$$

where ln denotes the natural logarithm. When λ is equal to 1, the series is analysed on its original scale, whereas the case $\lambda = 0$ corresponds to the logarithmic transformation. Other important special cases arise for fractional values of λ, e.g. the square root transformation ($\lambda = 1/2$). Obviously, for the transformation to be applicable, the series has to be strictly positive.

Suppose that the optimal forecast of the Box–Cox transformed series is denoted by $\tilde{y}_{t+h|t}(\lambda),\ h = 1, 2, \dots,$ where h is the forecast lead. Here, optimality is intended in the mean square error sense, so that $\tilde{y}_{t+h|t}(\lambda) = \mathrm{E}[y_{t+h}(\lambda)|\mathcal{F}_t]$ is the conditional mean of $y_{t+h}(\lambda)$, given the information set at time t, here denoted as $\mathcal{F}_t$. The conditional mean is typically available in closed form. Finally, let $\sigma_h^2(\lambda) = \mathrm{E}\{[y_{t+h}(\lambda) - \tilde{y}_{t+h|t}(\lambda)]^2|\mathcal{F}_t\}$ denote the h-step-ahead prediction error variance, which we assume, for the sake of simplicity, to be time-invariant.