MPRA

Munich Personal RePEc Archive

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

Tommaso, Proietti and Helmut, Luetkepohl

University of Sydney, European University Institute,
University of Rome Tor Vergata - Sefemeq department

18. July 2011

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

Tommaso Proietti[*]

University of Sydney and

Università di Roma "Tor Vergata"

Helmut Lütkepohl[†]

Department of Economics

European University Institute

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 one fifth of the series considered the Box-Cox transformation produces forecasts significantly better than the untransformed data at one-step-ahead horizon; in most of the 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 leads. We also discuss whether the preliminary in-sample frequency domain assessment conducted provides a reliable guidance which series should be transformed for improving significantly the predictive performance.

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

---

[*]Address for Correspondence: Merewether Building (H04), Discipline of Operations Management and Econometrics, The University of Sydney, NSW 2006, email: t.proietti@econ.usyd.edu.au.

[†]Address for Correspondence: Department of Economics, European University Institute, Via della Piazzola 43, I-50133 Firenze, Italy, email: helmut.luetkepohl@eui.eu

1   Introduction

Transformations aim at improving 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 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 simple structure, normal errors and constant error variance, it was subsequently modified by Box and Cox (1964).

The objective of this paper is assessing whether transforming a variable leads to an improvement in forecasting accuracy. The issue has already been debated in the time series literature. The use of the Box-Cox transformation as a preliminary specification step to 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).

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 post-sample evaluation, they concluded that the Box-Cox transformation does not lead to an improvement of the forecasting performance. Another important conclusion, supported also 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 is that the conditional expectation underlying the optimal forecast assumes that the transformed series is normally distributed. This assumption may not be realistic, however. 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 improved forecasting accuracy over 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 (2011a) conclude that using logs can be quite beneficial for forecasting. They also point out that there does not appear to be a reliable criterion for deciding between logs and levels for the purpose of maximizing forecast accuracy.

From the theoretical standpoint, Granger and Newbold (1976) provided a general analytical approach, based on the Hermite polynomials series expansion, to forecasting transformed series. Analytic 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. More recently, Pascual, Romo and Ruiz (2005) have proposed a bootstrap procedure for constructing prediction intervals for a series when an ARIMA model is fitted to its power transformation.

Finally, the Box-Cox transformation is popular in financial time series analysis and has been considered, for example, for forecasting volatility (see e.g. Higgins and Bera, 1992, and Goncalves and Meddahi, 2011) and price durations (Fernandes and Grammig, 2006).

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 in deciding whether or not to use the power transformation if forecasting is the objective. Our procedure has the advantage that it does not require normality assumptions which would be used in maximum likelihood procedures. Hence, it circumvents the problem noticed 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 mean square forecast error. We find that there is a certain percentage of series were 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 the following. In Section 2, after reviewing the Box-Cox transformation, we discuss the predictors of interest. In Section 3 we present the nonparametric procedure

3

for estimating the p.e.v. and the transformation parameter. Section 4 discusses the estimation results on the dataset. In Section 5 we judge the relevance of the transformation for out-of-sample forecasting by conducting a rolling forecasting experiment. Conclusions are drawn in Section 6.

2   Forecasting Box-Cox transformed series

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

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

where ln denotes the natural logarithm. When λ is equal to 1, the series is analysed in 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 transform ($\lambda = 1/2$). Obviously, for the transformation to be applicable, the series has to be strictly positive.

Suppose the optimal forecast of the Box-Cox transformed series is denoted by $\tilde{y}_{t+h|t}(\lambda)$, $h = 1, 2, \ldots$, 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 for simplicity we assume time-invariant.

We now consider the prediction of y_t on its original scale of measurement. The naïve forecast is obtained as the inverse Box-Cox transformation,

$$
\hat{y}_{t+h|t} = \begin{cases} (1 + \lambda \tilde{y}_{t+h|t}(\lambda))^{1/\lambda}, & \lambda \neq 0, \\ \exp(\tilde{y}_{t+h|t}(\lambda)), & \lambda = 0. \end{cases}
$$

This quantity corresponds to the median of the predictive distribution and, hence, it provides the minimum absolute error predictor.

The optimal predictor of y_{t+h} (i.e. its conditional expectation given the past), denoted by $\tilde{y}_{t+h|t}$, is

$$
\tilde{y}_{t+h|t} = \frac{1}{\sigma_h(\lambda)\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{y - \tilde{y}_{t+h|t}(\lambda)}{\sigma_h(\lambda)}\right)^2\right] (\lambda y - 1)^{1/\lambda} \mathrm{d}y \tag{2}
$$

Table 1: Optimal predictors of original variable for different Box-Cox transformation parameters

λ	$\tilde{y}_{t+h\|t}$
0	$\hat{y}_{t+h\|t} \exp\left(\sigma_h^2(\lambda)/2\right)$
$1/2$	$\hat{y}_{t+h\|t}\left(1 + \frac{1}{4}\frac{\sigma_h^2(\lambda)}{\hat{y}_{t+h\|t}}\right)$
$1/3$	$\hat{y}_{t+h\|t}\left(1 + \frac{1}{3}\frac{\sigma_h^2(\lambda)}{\hat{y}_{t+h\|t}^{2/3}}\right)$
$1/4$	$\hat{y}_{t+h\|t}\left(1 + \frac{3}{8}\frac{\sigma_h^2(\lambda)}{\hat{y}_{t+h\|t}^{1/2}} + \frac{3}{256}\frac{\sigma_h^4(\lambda)}{\hat{y}_{t+h\|t}}\right)$

if the transformed series is normally distributed (see Nelson and Granger, 1979). In general a closed form expression for this integral is not available. However, if $1/\lambda$ is a positive integer, Pankratz and Dudley (1987) and Proietti and Riani (2009) provide a closed form expression. Table 1 presents the expressions of the optimal predictors as functions of the naïve predictor $\hat{y}_{t+h|t}$ for selected values of λ.

3 Deciding on the Box-Cox transformation for prediction

The Box-Cox transformation parameter is usually estimated by maximum likelihood, assuming a parametric model for $y_t(\lambda)$; the parameter λ can be concentrated out of the likelihood, which is corrected by the Jacobian so as to take into account the change of scale of the observations. This approach is plausible if the Box-Cox transformation converts the distribution to a normal. Unfortunately, the results by Nelson and Granger (1979) indicate that the normality assumption for the transformed series may be problematic. Moreover, even though the forecasts are typically based on a parametric model, there is usually uncertainty regarding the right model. Therefore, we propose a nonparametric approach, according to which the transformation parameter is estimated as the value for which the prediction error variance (p.e.v.) of the series (after a normalization by the Jacobian of the transformation), is a minimum.

Our procedure is based on the normalized Box-Cox (NBC) transformation which is obtained by dividing $y_t(\lambda)$ by $\sqrt[n]{J}$, where $J = \prod_t \left|\frac{\partial y_t(\lambda)}{\partial y_t}\right|$ is the Jacobian of the transformation, which is equal to $g_y^{\lambda-1}$ with $g_y = [\prod_t y_t]^{1/n}$ being the geometric average of the original observations (Atkinson, 1973). This yields

$$z_t(\lambda) = g_y^{1-\lambda}y_t(\lambda). \tag{3}$$

This normalization is relevant if the aim is minimizing the one-step-ahead p.e.v. across the different values of λ. Notice that when $\lambda = 1$, the normalizing factor is unity, and the normalized transform sets the scale equal to that of the original observations.

We assume that $z_t = z_t(1)$ can be made stationary by differencing, that is, there exists a stationary representation $u_t = \Delta(L)z_t$, $t = 1, \ldots, n$, where $\Delta(L)$ is a polynomial in the lag operator, L, e.g. $\Delta(L) = (1 - L)^d$ or $\Delta(L) = (1 - L)(1 - L^s)$, for seasonal time series with seasonal period s. Obviously, if z_t is stationary, $\Delta(L) = 1$. Notice also that n has been reset so as to denote the number of observations available for u_t.

We estimate the transformation parameter by minimizing the p.e.v.. Notice that the one-step-ahead p.e.v. for z_t is the same as that of u_t, since $u_t - \mathrm{E}(u_t|\mathcal{F}_{t-1}) = z_t - \mathrm{E}(z_t|\mathcal{F}_{t-1})$.

If we let $f(\omega)$ denote the spectral density of u_t, and assume $\int_{-\pi}^{\pi} \ln f(\omega)d\omega > -\infty$, the one-step-ahead p.e.v. is defined, according to the usual Szegö-Kolmogorov formula, as the geometric average of the spectral density:

$$\sigma^2 = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln 2\pi f(\omega)d\omega\right].$$

The p.e.v. can be estimated nonparametrically by a bias-corrected geometric average of the periodogram. Letting $\omega_j = \frac{2\pi j}{n}$, $j = 1, \ldots, [n/2]$, denote the Fourier frequencies, where $[\cdot]$ is the integer part of the argument, the sample spectrum is defined as

$$I(\omega_j) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n}(u_t - \bar{u})e^{-\imath\omega_j t}\right|^2,$$

$\bar{u} = \frac{1}{n}\sum_t u_t$ and $\imath$ is the imaginary unit. Letting n^* denote $n/2 - 1$, if n is even, and $(n-1)/2$, if n is odd, Davis and Jones (1968) proposed the following estimator:

$$\hat{\sigma}^2 = \exp\left[\frac{1}{n^*}\sum_{j=1}^{n^*} \ln 2\pi I(\omega_j) + \gamma\right], \tag{4}$$

where $\gamma = 0.57722$ is Euler's constant.

Hannan and Nicholls (1977, HN henceforth) proposed replacing the raw periodogram ordinates by their non-overlapping averages of m consecutive ordinates,

$$\hat{\sigma}^2(m) = m\exp\left[\frac{1}{M}\sum_{j=0}^{M-1} \ln\left(\sum_{k=1}^{m} 2\pi I(\omega_{jm+k})\right) - \psi(m)\right]. \tag{5}$$

where $M = [(n-1)/(2m)]$ and $\psi(m)$ is the digamma function. The estimator (4) is obtained in the case $m = 1$. The large sample distributions of (5) and $\ln \hat{\sigma}^2(m)$ are, respectively,

$$\hat{\sigma}^2(m) \overset{a}{\sim} \text{N}\left(\sigma^2, \frac{2\sigma^4 m\psi'(m)}{n}\right), \quad \ln \hat{\sigma}^2(m) \overset{a}{\sim} \text{N}\left(\sigma^2, \frac{2m\psi'(m)}{n}\right).$$

The estimation of the optimal transformation parameter is carried out by a grid search over the λ values in the range (a, b), where typically, $b = -a = 2$. For each value of λ in the selected range the NBC transformation of the series, $z_t(\lambda)$, is computed according to (3), the stationarity transformation is obtained as $u_t(\lambda) = \Delta(L)z_t(\lambda)$ and the HN estimator (5) is computed. The value of λ that yields the minimum p.e.v. is the required estimate. Notice that a crucial assumption is that the stationarity inducing transformation, $\Delta(L)$, does not vary with λ, which is appropriate for the NBC.

When explanatory variables are present, such as trading days and Easter regressors for modeling calendar effects (see Cleveland and Devlin, 1982), interventions and seasonal dummy variables, the p.e.v. can be estimated from the frequency domain regression residual periodogram, as in Cameron (1978). Alternatively, we could use a weighted estimate of the p.e.v. based on a similar idea to band spectral regression (Engle, 1974), that puts a zero weight to the sample spectrum ordinates around the trading days and seasonal frequencies.

The latter may also be advocated as a more general strategy aiming at robustifying the non-parametric estimator of the p.e.v., by excluding some periodogram ordinates that could be affected by the stationarity inducing transformation. For instance, if $\Delta z_t = z_t - z_{t-1}$ is analyzed, leaving out the seasonal frequency may be thought of as a way of eliminating a deterministic seasonal component from the series. If we focus on $\Delta_s z_t = z_t - z_{t-s}$, instead, then the periodogram at the seasonal frequencies may get close to zero, so that the seasonal frequencies will contribute strongly and negatively to the p.e.v. estimate.

4   Estimation results

We apply the estimation method to a dataset consisting of 530 monthly time series, seasonally unadjusted, 379 of which are related to the index of industrial production (IPI) and 151 to the index of retail turnover for some Euro area countries, the UK and the US. For the IPI we consider series from Sectors B (Mining and quarrying), C (Manufacturing), D (Energy), and B–D, and

the series for the manufacturing sectors are from those identified by two digits of the NACE statistical classifications of economic activities (Sectors C1-C31). For the US we consider the 63 series for Market and Industry Group and the 32 series for Special Aggregates and Selected Detail (see `http://www.federalreserve.gov/releases/g17/table1_2.htm` for more details). For retail turnover, we focus on the series available with code starting with G47 (Retail trade, except of motor vehicles and motorcycles). The series were obtained from Eurostat (`http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/`), the Federal Reserve and the US Census Bureau. The breakdown of the series by country and their sample period is available in Table 2.

The first objective of our analysis is to check if our estimation method suggests that transforming our series is useful for reducing the p.e.v. and to narrow down the range of λ values to be considered. Given the previous results, e.g. by Nelson and Granger (1979), one may expect the logarithmic transformation ($\lambda = 0$) to be of particular importance. For each individual time series the transformation parameter was estimated as the minimizer of the one-step-ahead p.e.v. of the NBC transform, which is computed by the HN nonparametric estimator (5) using $m = 3$. This particular choice for the value of m was suggested by a Monte Carlo simulation experiment, not reported for brevity, according to which, for the sample sizes considered, setting $m = 3$ provides the most reasonable compromise between bias and variance in estimating the p.e.v., the estimator with $m = 1$ (the Davies and Jones estimator) being characterised by high sampling variance; on the contrary, larger values for m lead to a minor reduction in the variance and larger biases. As the series are strongly seasonal, we assume that $u_t = z_t - z_{t-12}$ is stationary. Before summarizing the results, it may be instructive to consider an example series for which our method suggested the need for a transformation.

Figure 1 displays the French industrial production series for the branch *Manufacture of wearing apparel* along with the interval estimates of the logarithmic p.e.v.. Clearly, the volatility of the series appears to be linked to its level, suggesting that a Box-Cox transformation can make it more homogeneous. The minimum p.e.v. occurs for $\lambda = 0.28$. The transformation indeed helps stabilizing the amplitude of the series, so that $z_t(\lambda) - z_{t-12}(\lambda)$ looks more like a covariance stationary series (not shown). In contrast, the yearly changes on the original scale, $y_t - y_{t-12}$, are clearly heteroscedastic, as suggested by Figure 1. An approximate 95% confidence interval for λ is (0.01, 0.61). The latter is computed as the smallest set of λ values which has an interval estimate at the same confidence level that includes the minimum p.e.v. estimate. Notice that the coverage may

Table 2: Breakdown of the time series analysed by country, sample period, number of time series.

Index of industrial production

Country	Sample period	Number of series
Austria	1996.1-2010.12	28
Belgium	1995.1-2010.12	27
Finland	1990.1-2010.12	20
France	1990.1-2010.12	28
Germany	1991.1-2010.12	28
Greece	2000.1-2010.12	29
Italy	1990.1-2010.12	27
Netherlands	1990.1-2010.12	22
Portugal	1995.1-2010.12	20
Spain	1980.1-2010.12	28
UK	1990.1-2010.12	28
US	1947.1-2010.12	94

Index of retail turnover

Country	Sample period	Number of series
Austria	1999.1-2010.12	6
Belgium	1998.1-2010.12	15
Finland	1995.1-2010.12	14
France	1994.1-2010.12	14
Germany	1994.1-2010.12	15
Greece	1995.1-2010.12	13
Italy	2000.1-2010.12	14
Netherlands	1996.1-2010.12	9
Portugal	1995.1-2010.12	9
Spain	2000.1-2010.12	14
UK	2000.1-2010.12	14
US	1992.1-2010.12	14

Figure 1: French index of industrial production for *Manufacture of wearing apparel*. Series and interval estimates of $\ln \sigma^2$ obtained by the HN estimator using $m = 3$.
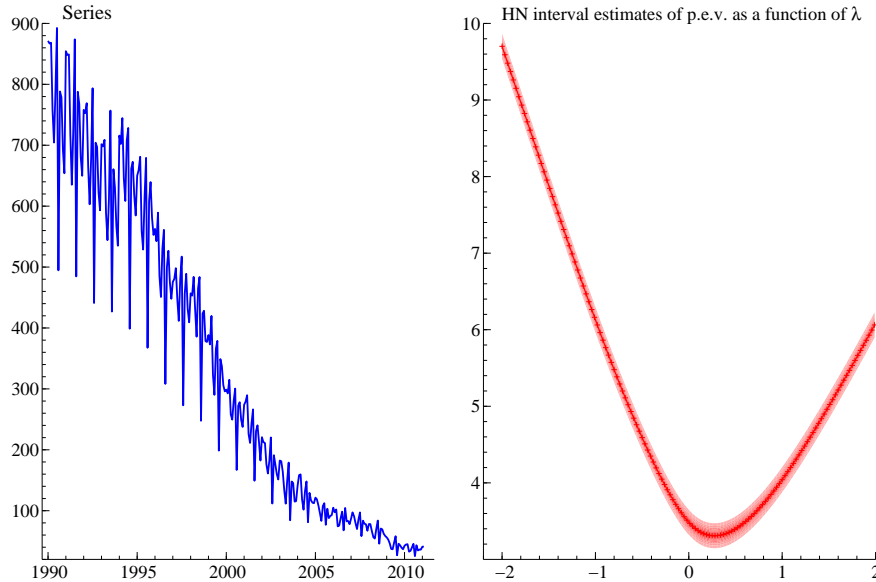

be less than the nominal one due to the fact that the interval is constructed using the asymptotic variance, whereas the finite sample variance can be larger. Hence, the logarithmic transformation ($\lambda = 0$) could also be adequate.

The distribution of the 530 point estimates of λ, displayed in Figure 2, is centered around the mean value 0.29, with a standard deviation of 0.64. The percentage of cases in which the value $\lambda = 1$ (i.e., no transformation needed) is not contained in the 95% confidence interval for the parameter is 18.68% (99 cases in total), of which only 5.05% (5 cases) resulted in λ estimates significantly greater than zero, whereas for the remaining 94 series the estimated value is not significantly different from zero. The main conclusion is that, if a transformation is indicated, in the vast majority of the cases it can be safely taken as the logarithmic transformation. The complete distribution can be considered as a mixture of two distributions, also plotted in Figure 2, the first arising when λ is not significantly different from 1, and the second in the contrary case. The latter is shifted to the left with mean -0.27 and standard deviation 0.44 so that $\lambda = 0$ is in its central region.

These results were obtained making no provision for calendar effects or for possible overdifferencing caused by the stationarity inducing transformation Δ_{12}. However, if we exclude from

Figure 2: Density of the nonparametric estimator of the transformation parameter, λ, estimated from 530 time series.
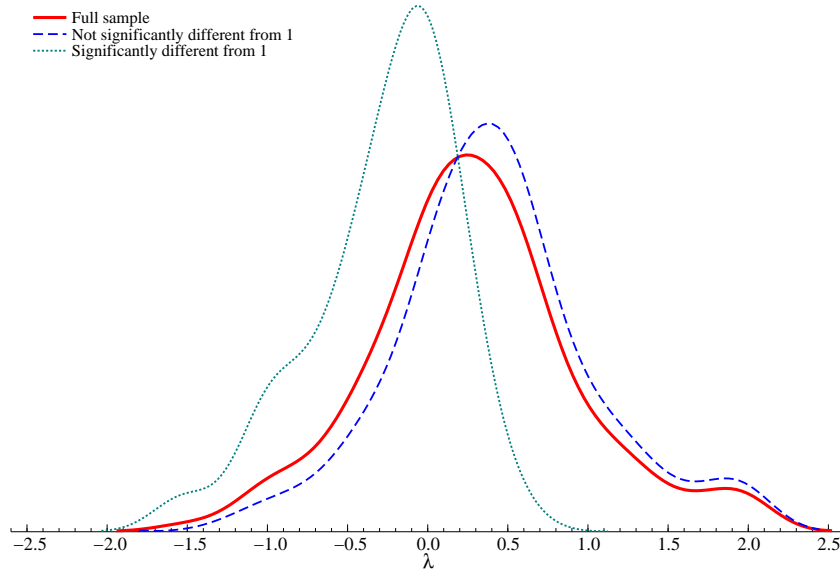
(5) the frequencies ω_j corresponding to trading days effects (the most prominent being around $0.348 \times 2\pi$ and $0.432 \times 2\pi$, see Cleveland and Devlin, 1980) as well as the seasonal frequencies $2\pi j/12, j = 1, \ldots, 6$, the estimates of λ are not affected, as the deletion results only in a vertical shift in the p.e.v. as a function of λ.

5   Empirical forecast comparison

This section aims at supporting the previous evidence concerning the need to transform the data by means of a genuine out-of-sample forecasting experiment. Our estimator and decision rule was based on all the available data and did not look at the relevance of the Box-Cox transformation for improving the out-of-sample forecasting ability. It is well known that the performance of a transformation model within the sample may not be coincident with that outside the available sample. Thus, in this section, using a simple though flexible benchmark model, we will compare the forecasting performance of different predictors arising when λ is fixed at specific values, including the one estimated nonparametrically.

Our rolling forecast experiment is designed as follows.

1. The size of the rolling window is fixed at 6 or 10 years of monthly observations depending on the length of the series (we use 10 years if the length is greater than 15 years).

2. For each rolling sample we fit an $AR(p)$ model to the yearly changes of the power transformation of the series, $\Delta_{12}y_t(\lambda)$, with regression effects,

$$\phi_p(L)\Delta_{12}y_t(\lambda) = \beta_0 + \beta'\Delta_{12}x_t + \epsilon_t, \quad \epsilon_t \sim \text{WN}(0, \sigma^2),$$

where x_t are 6 trading days regressors and an Easter variable, that account for calendar effects, and $y_t(\lambda)$ is the Box-Cox transform of the series. Along with the nonparametric estimate of λ, $\hat{\lambda}$, we consider also the following values of λ: $\lambda = 0$ (log transformation), $\lambda = 1/3$, $\lambda = 1/2$ and $\lambda = 1$ (no transformation), which represent the most relevant cases for which the out-of-sample forecasts are available in closed form. When the transformation parameter is $\hat{\lambda}$, we use the naïve predictor, as we do not want to rely on normality or other distributional assumptions when this transformation is considered.

3. The order of the autoregression is selected according to the Schwarz information criterion and the model is estimated by least squares. The maximum lag order considered is 12. We leave open the possibility that the orders of the automatically identified AR models differ with λ (Granger and Newbold, 1976).

4. Conditionally on the estimated model we compute the forecasts of the original levels y_t and the yearly growth rates, $g_t = \frac{\Delta_{12}y_t}{y_{t-12}}$, for all forecast horizons up to $H = 24$ steps ahead and compare them to the observed values. For $\lambda = 0, 1/2$ and $1/3$ we compute both the optimal and the naïve predictor according to the formulae presented in Section 2. Hence, we consider a total of eight competing forecasts (including $\lambda = 1$ and $\hat{\lambda}$).

5. The rolling window is then moved one month forward and the steps 2–4 are iterated until we reach observation $n - H$.

6. We summarise the distribution of the prediction errors at different horizons using the mean square forecast error. Subsequently, we select the best performing predictor at horizon h among the seven predictors obtained for $\lambda = 0, 1/2$ and $1/3$, and $\lambda = \hat{\lambda}$ and test the equality of the prediction mean squared errors with those computed for the benchmark model ($\lambda = 1$), using the Harvey, Leybourne and Newbold (1997) version of the Diebold-Mariano

(Diebold and Mariano, 1995, DM henceforth) test statistic.[1] Significant values suggest that transforming the data helps reducing the forecast mean square error.

Table 3 summarises the empirical results by presenting the number of times (over the 530 series analysed) in which the best performing predictor with $\lambda \neq 1$, listed in the rows of the table, outperformed significantly the benchmark predictor, computed on the untransformed series. Significant differences are measured by the modified DM test using a 5% significance level.

As far as one-step-ahead forecasting is concerned only in 18% of the cases the Box-Cox predictor using $\lambda \neq 1$ provides systematically better predictions than the benchmark, for which the data are not transformed. The percentage is slightly higher (20%) if we aim at predicting growth rates. Secondly, as the horizon h increases, the case for transforming the data is less strong: if the forecast horizon is $h = 12$, i.e. a year ahead, the proportion of cases in which the benchmark is outperformed significantly at the same level reduces to a mere 7% for the levels and 8% for the growth rates. These numbers are not much larger than the significance level and, hence, what one would expect if the null hypothesis was true. Thus, for longer forecast horizons there is very little evidence that the Box-Cox transformation can improve forecast efficiency.

In the cases for which a transformation is relevant, the nonparametric estimator of λ that we have proposed, $\hat{\lambda}$, provides the best predictor in a percentage of cases varying with h. For the levels it varies between 28% and 39% whereas the range for the growth rates is from 18% to 38%. The

---

[1]Letting $e_{\lambda,j}$, and $e_{1,j}$, $j = 1, \ldots, J$, denote the sequence of h-step-ahead forecast errors, respectively for the best predictor using $\lambda = 0, 1/3, 1/2, \hat{\lambda}$ and the benchmark predictor ($\lambda = 1$), and defining the quadratic loss differential sequence $d_j(h) = e_{\lambda,j}^2 - e_{1,j}^2$, $j = 1, \ldots, J$, the DM test of the null hypothesis of equal forecast accuracy, $H_0 : \mathrm{E}(d_j(h)) = 0$, versus the one sided alternative that the model with $\lambda \neq 1$ provides more accurate forecasts, $H_1 : \mathrm{E}(d_j(h)) < 0$, is based upon the statistic

$$DM(h) = \frac{\bar{d}(h)}{\sqrt{\sigma_d^2}}, \quad \bar{d}(h) = \frac{1}{J} \sum_j d_j(h), \quad \sigma_d^2 = \frac{1}{J} \left[c_0 + 2 \sum_{k=1}^{q-1} \frac{J-k}{J} c_k \right],$$

where c_k is the sample autocovariance of $d_j(h)$ at lag k and σ_d^2 is a consistent estimate of the variance of the loss differential. In our applications the value of the truncation parameter is set equal to $q = \max(h, 4)$ and we use the DM statistic with the small sample modification proposed by Harvey, Leybourne and Newbold (1997) which provides a correction for the bias of σ_d^2 as an estimator of the variance of $d_j(h)$:

$$DM^*(h) = DM(h) \left[\frac{J + 1 - 2q + q(q+1)/J}{J} \right]^{1/2}.$$

Under the null hypothesis the reference distribution is Student's t with $J - 1$ degrees of freedom, denoted T_{J-1}.

logarithmic transformation overall has the best performance, although there may be cases in which the estimated λ is not significantly different from 0. Notice also that, somewhat counter-intuitively, the naïve predictor outperforms the optimal one at short forecast leads. This result is in line with observations by Bårdsen and Lütkepohl (2011) in a comparison of forecasts based on levels and logs in multivariate systems. They explain the phenomenon by the estimation uncertainty involved in computing the optimal forecast. An alternative explanation may, of course, be that the normality assumption underlying the optimal predictor is invalid. The square root transformation plays only a negligible role.

Given that there are some series for which forecast efficiency gains can be obtained with the Box-Cox transformation for short horizons, it would be useful to validate the nonparametric method as a tool for deciding the need for a transformation.

As far as one-step-ahead prediction of the levels is concerned, setting $\lambda = \hat{\lambda}$ yields an improvement in the forecasting performance (i.e. a reduction in the mean square forecast error) with respect to $\lambda = 1$ in 54% of the cases (288 out of 530). However, the improvement is statistically significant only in 18% of the cases, a proportion in line with the result in Table 3.

To compare the results of nonparametric estimation with the out-of-sample evidence we categorize the nonparametric estimates $\hat{\lambda}$ in three groups: those which were not significantly different from 1, suggesting to leave the series untransformed ($\lambda = 1$), those which were significantly different from 1, but not different from zero, suggesting the log transformation ($\lambda = 0$), and finally those that were significantly different from both 0 and 1 ($0 < \lambda < 1$). Table 4 cross-tabulates the nonparametric estimation results with the results of the rolling forecasting experiment, which categorize the series in two groups, according to whether the modified DM test was significant or not at the 5% level. Pearson's χ^2-test of independence resulted in a value of 9.83 with p-value 0.007 and, hence, suggests that the results are not independent. In other words, there is a clear association between our procedure suggesting a transformation and getting significantly better forecasts by transforming the series. Moreover, in 72% of the cases the nonparametric method that we propose provided a reliable guidance for the outcome of the out-of-sample rolling forecasting exercise. The number of false positive (70) and false negative decisions (69), jointly representing 28% of the cases, might be explained by the different nature of the two methods. While the nonparametric method looks at the evidence for a transformation within the sample and rests crucially on the assumption of stationarity of the underlying process, the rolling forecast experiment evaluates the

Table 3: Rolling forecast experiment: number of series for which the predictor listed in the rows resulted as the best predictor and outperformed significantly the benchmark ($\lambda = 1$) at the 5% level according to the one-sided modified DM test. The percentages in the last row are obtained by dividing the totals in the previous row by 530.

Levels

λ	Forecast horizons					
	$h=1$	$h=2$	$h=3$	$h=6$	$h=12$	$h=24$
$\hat{\lambda}$	35	30	28	13	12	16
0 (optimal)	10	9	8	13	17	14
0 (naïve)	34	34	24	15	5	7
1/3 (optimal)	2	2	4	2	1	2
1/3 (naïve)	13	7	5	2	1	1
1/2 (optimal)	2	1	2	0	0	0
1/2 (naïve)	1	1	1	1	1	1
Total	97	84	72	46	37	41
Percentage	18.30	15.85	13.58	8.68	6.98	7.74

Growth rates

λ	Forecast horizons					
	$h=1$	$h=2$	$h=3$	$h=6$	$h=12$	$h=24$
$\hat{\lambda}$	31	29	27	18	8	8
0 (optimal)	26	24	17	26	27	26
0 (naïve)	31	25	16	11	6	5
1/3 (optimal)	11	2	5	3	3	0
1/3 (naïve)	4	3	2	0	0	0
1/2 (optimal)	1	0	0	1	0	0
1/2 (naïve)	1	0	0	1	0	0
Total	106	86	72	61	44	39
Percentage	20.00	16.23	13.58	11.51	8.30	7.36

Table 4: Comparison of results of HN estimator and rolling forecastas

| Nonparametric Estimator	Modified DM test		Total
	Significant	Not Significant	
$\lambda = 0$	27	67	94
$0 < \lambda < 1$	2	3	5
$\lambda = 1$	69	362	431
Total	98	432	530

performance out-of-sample and can accommodate time variation and local nonstationarities. Moreover, note that a forecast based on a transformed series may have a smaller mean square forecast error without being significantly superior to the benchmark.

In addition, Figure 3 displays the empirical distribution function of the p-values of the statistic $DM^*(1)$, $P(T_{J_i-1} < DM_i^*(1))$, $i = 1, \ldots, 530$, respectively for the 99 cases in which the nonparametric estimator suggested to transform the data ($\lambda \neq 1$) and contrasts it with the empirical distribution function of the 431 cases for which a transformation was not suggested. The two distributions are different and in fact the Kolmogorov-Smirnov test of equality of the two distributions versus the alternative that the former is larger than the latter, based on the maximum distance between the two distribution functions, takes the value 0.185, with asymptotic p-value 0.005. This is a further confirmation that the preliminary nonparametric assessment of the need to transform the series can be useful. These results are quite promising, especially when we compare them to related results in the literature. For example, Lütkepohl and Xu (2011, 2011a) consider the simpler setting of comparing forecasts based on levels and logs only and conclude that they did not find a reliable decision rule for choosing between the two possibilities. Thus, our method appears to be a valuable tool.

6   Conclusions

It is argued that previous studies of the Box-Cox transformation as a means for improving forecast accuracy may be distorted by the assumption of a normally distributed transformed series. The latter assumption is often adopted in estimating the transformation parameter and computing the

Figure 3: Empirical distribution functions of the p-values of the modified DM test at horizon $h = 1$ for the two subpopulations consisting respectively of the 99 cases in which the nonparametric estimator suggested $\lambda \neq 1$ and of the 431 remaining cases for which $\lambda = 1$.
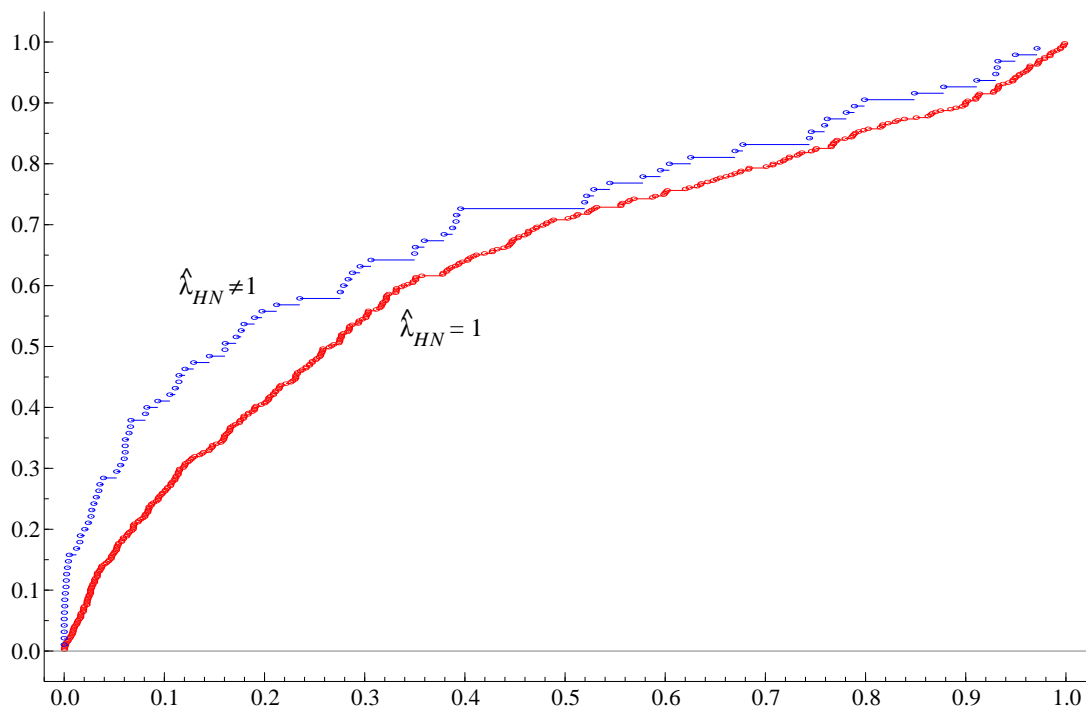
optimal forecast of the original series. We propose a distribution free, nonparametric method for estimating the Box-Cox transformation parameter and perform a large scale forecast comparison based on a much larger set of time series than previous studies. More precisely we consider a set of 530 monthly, seasonal time series related to industrial production and retail turnover of a large number of countries. We find three main results.

First, in only about 20% of the cases, a 95% confidence interval around the estimated transformation parameter does not contain the value of one which corresponds to no transformation. Thus, using a transformation parameter significantly different from one as a criterion for considering the Box-Cox transformation, for roughly 20% of the series a transformation is indicated. Clearly, if in one fifth of the series forecast improvements are possible, this is too large a set of series to be ignored if one is seriously interested in improving forecast accuracy. Out of the series for which transformations are indicated, only a very small fraction (about 5%) has a parameter significantly different from 0 which corresponds to the log transformation. Thus, in most cases where a transformation is indicated a log transformation may be a good choice. This result is well in line with Nelson and Granger (1979) who perform a forecast comparison for a much smaller set of time series.

As a second main result, our forecast comparison shows that transformations can indeed improve forecast accuracy at short horizons. For about 20% of the series, the one-step-ahead forecasts are significantly improved by using a Box-Cox transformation. For longer-term forecasts the advantage of the transformation diminishes, however. It turns out that, although the log transformation is indeed very successful in providing the best forecasts when a transformation is needed, estimating the transformation parameter by our method results in the best forecasts in about one third of the cases where a significant improvement is found, at least for short horizons. Moreover, the naïve predictor obtained by just inverting the transformation performs overall better than the optimal predictor which is based on the conditional expectation and uses normality assumptions. Again the latter result conforms well with previous related studies.

Our third main result is that our nonparametric approach to estimating the p.e.v. can provide good guidance and be a useful substitute for a computationally more demanding rolling forecast exercise for deciding on the need for a Box-Cox transformation. In general, when the approximate 95% nonparametric interval estimate does not cover $\lambda = 1$, it is also the case that the rolling forecast experiment points at the need to transform the series for getting significantly better forecasts (judging significance by a modified DM test). The number of false positive and false negative

18

decisions is typically small. The false decisions may be due to structural change and inhomogeneity of the series.

References

Atkinson, A. C. (1973). Testing transformations to normality, *Journal of the Royal Statistical Society B*, 35, 473–479.

Bårdsen, G. and Lütkepohl, H. (2011). Forecasting levels of log variables in vector autoregressions, *International Journal of Forecasting*, forthcoming.

Bell, W. R. and S. C. Hillmer (1983). Modeling time series with calendar variation. *Journal of the American Statistical Association*, 78, 526–534.

Box G. E. P. and Cox D.R. (1964). An analysis of transformations (with discussion), *Journal of the Royal Statistical Society B*, 26, 211–246.

Box, G.E.P. and Cox D.R. (1982). An analysis of transformation revisited, rebutted, *Journal of the American Statistical Association*, 77, 209–210.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Box G. E. P. and Jenkins, G.M. (1973). Some comments on a paper by Chatfield and Prothero and on a review by Kendall, *Journal of the Royal Statistical Society A*, 136, 337-352.

Cameron, M.A. (1978). The prediction variance and related statistics for stationary time series, *Biometrika*, 65, 283–296

Chatfield C., and Prothero, D. L. (1973). Box-Jenkins seasonal forecasting: problems in a case-study, *Journal of the Royal Statistical Society A*, 136, 295–315.

Cleveland, W.S. and Devlin, S.J. (1982). Calendar effects in monthly time series: modeling and adjustment, *Journal of the American Statistical Association,* 77, 520–528.

Davis, H.T. and Jones, R.H. (1968). Estimation of the innovation variance of a stationary time series, *Journal of the American Statistical Association*, 63, 141–149.

Diebold, F.X. and Kilian, L. (2001). Measuring predictability: theory and macroeconomic applications, *Journal of Applied Econometrics*, 16, 657–669.

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

Engle, R.F. (1974). Band spectrum regression, *International Economic Review*, 15, 1–11.

Fernandes, M. and Grammig J. (2006). A family of autoregressive conditional duration models, *Journal of Econometrics*, 130, 1–23.

Goncalves, S. and Meddahi, N. (2011). Box-Cox transforms for realized volatility, *Journal of Econometrics*, 160, 129–144.

Granger, C.W.J. and Newbold, P. (1976). Forecasting transformed time series, *Journal of the Royal Statistical Society B*, 38, 189-203.

Guerrero, V. M. (1993). Time-series analysis supported by power transformations, *Journal of Forecasting*, 12, 37–48.

Hannan, E.J., and Nicholls, D.F. (1977). The estimation of prediction error variance, *Journal of the American Statistical Association*, 72, 834-840.

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

Higgins, M.L. and A. Bera (1992). A class of nonlinear ARCH models. *International Economic Review*, 33, 137-158.

Hopwood, W.S., McKeown, J.C. and Newbold P. (1981). Power transformations in time-series models of quarterly earnings per share, *Accounting Review*, 56, 927–933.

Lütkepohl, H. and Xu, F. (2011). Forecasting annual inflation with seasonal monthly data: using levels versus logs of the underlying price index, *Journal of Time Series Econometrics*, 3(1), article 7.

Lütkepohl, H. and Xu, F. (2011a). The role of the log transformation in forecasting economic variables, *Empirical Economics*, forthcoming.

Nelson, C.R. (1976). The interpretation of R^2 in autoregressive-moving average time series models, *American Statistician*, 30, 175–180.

Nelson, H.L. and Granger, C.W.J. (1979). Experience with using the Box-Cox transformation when forecasting economic time series, *Journal of Econometrics*, 1979, 57–69.

Neyman J.N. and Scott E.L. (1960). Correction for bias introduced by a transformation of variables, *Annals of Mathematical Statistics*, 31, 3, pp. 643–655.

Pankratz, A. and Dudley, U. (1987). Forecasts of power-transformed series, *Journal of Forecasting*, 6, 239–248.

Pascual, L., Romo, J. and Ruiz, E. (2005). Bootstrap prediction intervals for power-transformed time series, *International Journal of Forecasting*, 21, 219–235.

Proietti, T. and Riani, M. (2009). Transformations and seasonal adjustment, *Journal of Time Series Analysis*, 30, 47–69.

Tukey, J. W. (1957) On the comparative anatomy of transformations, *Annals of Mathematical Statistics*, 28, 602-632.

Tunnicliffe Wilson, G. (1973). Discussion of the paper by Dr Chatfield and Dr Prothero, *Journal of the Royal Statistical Society A*, 136, 315–319.