Economics Letters 129 (2015) 112–115

Contents lists available at ScienceDirect

Economics Letters

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

Prediction bias correction for dynamic term structure models

Eran Raviv *

Econometric Institute, Erasmus University Rotterdam, The Netherlands
Tinbergen Institute, The Netherlands

HIGHLIGHTS

- Residuals from a NS model do not adhere to the familiar white noise assumption.
- An adjustment is proposed and backtested with a focus on forecasting performance.
- Large improvement in forecasting power is achieved over time and across yields.
- Gains are robust to different time periods and to different model specifications.

ARTICLE INFO

Article history:
Received 14 September 2014
Received in revised form
17 January 2015
Accepted 19 January 2015
Available online 11 February 2015

JEL classification:
E43
E47
G17

Keywords:
Yield curve
Nelson–Siegel
Time varying loadings
Factor models

ABSTRACT

When the yield curve is modelled using an affine factor model, residuals may still contain relevant information and do not adhere to the familiar white noise assumption. This paper proposes a pragmatic way to improve out of sample performance for yield curve forecasting. The proposed adjustment is illustrated via a pseudo out-of-sample forecasting exercise implementing the widely used Dynamic Nelson–Siegel model. Large improvement in forecasting performance is achieved throughout the curve for different forecasting horizons. Results are robust to different time periods, as well as to different model specifications.

1. Introduction

The yield curve is key statistic for the state of the economy, widely tracked by both policy makers and market participants. Accurate prediction of the curve is of great use for investment decision, risk management, derivative pricing and inflation targeting. It is therefore no surprise to witness the vast literature related to the modelling and forecasting of the term structure.

Notable landmarks are the early work of Vasicek (1977) and Cox et al. (1985) through Duffie and Kan (1996) and Dai and Singleton (2002), all of which focus on the class of affine term structure models, and Hull and White (1990) and Heath et al. (1992), who focus on fitting the term structure under no arbitrage restrictions.

A popular choice for a prediction model is the one put forward by Diebold and Li (2006) (henceforth DL). They successfully demonstrate how a variant of the Nelson–Siegel model (Nelson and Siegel, 1987) can be used for prediction. The model itself is essentially a common parametric function, which is flexible enough to describe the many possible shapes assumed by the yield curve. In their seminal paper from 2006, DL build a dynamic framework for the entire yield curve, a dynamic Nelson–Siegel model (henceforth NS). Factors are estimated recursively using standard cross-sectional OLS, and evolve according to an AR(1) process. This approach has at least two appealing aspects. First, time-varying parameters can be easily interpreted as the well-known triplet level, slope and curvature. These three latent factors have been shown to be the driving force behind the yields co-movement (Litterman and Scheinkman, 1991). Second, estimation is easy and robust, analytical solution is at the ready which makes recursive estimation simple and fast. This is in stark contrast to a Maximum-Likelihood estimation which despite being theoretically more efficient (conditional on normality assumptions) is prohibitively compu-

* Correspondence to: APG Asset management, Gustav Mahler Square 3, 1082 MS Amsterdam, The Netherlands.
E-mail address: eran.raviv@apg-am.nl.

http://dx.doi.org/10.1016/j.econlet.2015.01.022
0165-1765/

tationally expensive, sensitive to starting values and sensitive to search algorithm used.

At the very heart of affine term structure models, lies the decomposition of the curve into the common part and the idiosyncratic part. When the yield curve is properly spanned by a small set of common factors, the idiosyncratic part can be treated as white noise. Specifically, there should be no autocorrelation or bias once underlying factors are accounted for. However, in practice, it may not be the case. Model errors may exhibit clear deviations from those assumptions. This issue has recently gained increased attention. Hamilton and Wu (2011) and Duffee (2011) document term structure model errors that exhibit high serial correlation. In terms of forecasting, Bauer et al. (2012) claim that parameters of a dynamic term structure model incur small-sample bias.

Usually, we target the yields themselves, factor modelling is a means to an end. Here we suggest a pragmatic way to correct for the effect brought about by this bias, working directly with out-of-sample model errors. Once errors deviate from the white noise assumption, a simple correction is applied to directly extract remains of information. As recently been suggested, conditional on the existence of such bias, this has the potential to improve forecasting performance.

We empirically illustrate this point using the NS model, but the procedure is valid for any factor model used. The NS model is compared favourably in terms of forecasting performance to other less parsimonious models (Mönch, 2008, for example). The model fits the curve well, however, the residuals from the fit *over time* exhibit (1) strong autocorrelation and (2) mean which significantly deviates from zero. These stylized facts can be exploited to improve prediction.

The next section motivates and presents the proposed adjustment. Section 3 presents the empirical results while Section 4 concludes. In an Appendix the interested reader can find results from other term structure model specifications which are considered as a robustness check.

2. The model and the proposed bias correction

For the yield curve of interest rates, using the well known latent factor model suggested by Nelson and Siegel (1988), the loadings are predetermined functions of maturity τ. The representation given by Diebold and Li (2006) to this model is given by:

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t}\left(\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}\right) + \beta_{3,t}\left(\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)\right) + \varepsilon_t, \tag{1}$$

where available maturities at time t, $\boldsymbol{\tau} = \{\tau_1, \dots, \tau_M\}$.

The parameter β_1 can be interpreted as the long-term interest rate, or a "leve" factor. The parameter β_2 determines how fast we the yield approaches its long term value, and is known as the "slope" factor. The parameter β_3 determines the size and shape of the hump, and is known as the "curvature" factor. Lastly, the parameter λ_t determines the decay rate for the loadings on the second factor, and the maturity at which loading on the third factor is maximized.[1] In the special case where $\lambda_t = \lambda\ \forall t$, the factors $\boldsymbol{\beta}_t$ are obtained using a simple cross sectional regression across available maturities at time t. The residuals $\boldsymbol{\varepsilon}_t = \{\varepsilon_{t,1}, \dots, \varepsilon_{t,M}\}$ are assumed to be white noise. Note that the assumption concern the cross sectional aspect of the model, and do not necessarily hold over time. To be more specific, the model does not assume residuals that are independent over time, nor that factor's volatility are constant over time. For example Hautsch and Yang (2012) show that by extracting time varying volatility components, mean forecasts may be similar yet sharper densities are produced, testifying to decreased forecast uncertainty.

The h-step-ahead prediction is given by:

$$\widehat{y}_{t+h}(\tau) = \widehat{\beta}_{1,t+h} + \widehat{\beta}_{2,t+h}\left(\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}\right) + \widehat{\beta}_{3,t+h}\left(\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)\right) \tag{2}$$

with

$$\widehat{\boldsymbol{\beta}}_{t+h} = \widehat{\boldsymbol{\alpha}} + \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\beta}}_t, \tag{3}$$

where $\widehat{\boldsymbol{\beta}}_t$ is a 3×1 vector, as is $\widehat{\boldsymbol{\alpha}}$. $\widehat{\boldsymbol{\Gamma}}$ is a 3×3 coefficient matrix which may or may not be diagonal. Arguments can be raised in favour and against a diagonal restricted $\widehat{\boldsymbol{\Gamma}}$ matrix. Diagonal restricted $\widehat{\boldsymbol{\Gamma}}$ has less parameters so less estimation uncertainty, more parameters may result in a noisier forecast. However, unrestricted $\widehat{\boldsymbol{\Gamma}}$ allows for conditional cross-correlation between factors which may be important. In the forecasting exercise we use a diagonal restricted $\boldsymbol{\Gamma}$ as advocated in Diebold and Li (2006). Results from the fully parametrized $\boldsymbol{\Gamma}$ are presented in subsequent section for completeness.

Now define the out-of-sample forecasting errors from the chosen forecasting model as:

$$e_{t+h}(\tau) = y_{t+h}(\tau) - \widehat{y}_{t+h}(\tau).$$

The mapping between the factors and the yields is done using cross sectional projection. Therefore there is a possibility that the residuals *over time*, still contain information to be exploited. The information can be in the form of errors which have non-zero mean or strong autocorrelation, features that can be observed even for the in-sample residuals.

A pragmatic way to extract potential remains of information is by using an AR model, so that the forecast for the out-of-sample error is obtained by:

$$\widehat{e}_{t+h}(\tau) = \delta(\tau, h) + \rho(\tau, h)e_t(\tau). \tag{4}$$

In this equation δ is interpreted as the bias of the forecast, and ρ is the autocorrelation coefficient. Keeping our focus on prediction, the adjusted forecast is given by:

$$\begin{aligned}\widehat{y}^{adj}_{t+h}(\tau) &= \widehat{y}_{t+h}(\tau) + \widehat{e}_{t+h}(\tau)\\ &= \widehat{y}_{t+h}(\tau) + \widehat{\delta}(\tau, h) + \widehat{\rho}(\tau, h)\widehat{e}_t(\tau).\end{aligned} \tag{5}$$

The parameters δ and ρ are estimated using a direct projection of the out-of-sample errors on their past, in the same manner that we determine the AR coefficients for factors dynamics. In essence, we extract potential information in model errors and use it to adjust our prediction for the next period.

3. Empirical results

In this section we describe the data, followed by estimation methods and forecasting results using our proposed adjustment.

3.1. Data description

We use the same data as in DL (2006), a balanced panel data of 17 maturities.[2] The last data point in their dataset is 12/2000.

[1] A more detailed description can be found in Diebold and Rudebusch (2013).

[2] The data can be downloaded from http://www.ssc.upenn.edu/~fdiebold/papers/paper49/FBFITTED.txt. A summary statistics table can be found in the Appendix along with a plot of the data and the NS-based factors.