

Raviv, Eran

Working Paper

Prediction Bias Correction for Dynamic Term Structure Models

Tinbergen Institute Discussion Paper, No. 13-041/III

Provided in Cooperation with:

Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Raviv, Eran (2013) : Prediction Bias Correction for Dynamic Term Structure Models, Tinbergen Institute Discussion Paper, No. 13-041/III, Tinbergen Institute, Amsterdam and Rotterdam,
<https://nbn-resolving.de/urn:nbn:nl:ui:15-1765/39191>

This Version is available at:

<https://hdl.handle.net/10419/87460>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

TI 2013-041 /III
Tinbergen Institute Discussion Paper



Prediction Bias Correction for Dynamic Term Structure Models

Eran Raviv

*Econometric Institute Erasmus School of Economics, Erasmus University Rotterdam and
Tinbergen Institute*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Prediction bias correction for dynamic term structure models

Eran Raviv

Econometric Institute, Erasmus University Rotterdam

Tinbergen Institute

December 2012

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.

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

JEL-code: E43; E47; G17

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, Ingersoll, and Ross (1985), through Duffie and Kan (1996) and Dai and Singleton (2000), all of which focus on the class of affine term structure models, and Hull and White (1990) and Heath, Jarrow, and Morton (1992), who focus on fitting the term structure under no arbitrage restrictions.

The value of no arbitrage restrictions is that, given the factors, we need only a small number of parameters to determine the shape of the curve. The number of cross-sectional free parameters is substantially larger without forcing no-arbitrage. So, if data agree with no-arbitrage restrictions, intuition tells us that imposing these restrictions will render better out-of-sample results.

Despite these important advances, Duffee (2002) documents that in the context of forecasting, good out-of-sample performance remains a challenge. More recently, Duffee (2011) finds that imposing no-arbitrage restrictions does not improve forecasting performance. Mönch (2008), De Pooter et al. (2010) and Exterkate et al. (2012) add macro variables and document that they contain predictive information for the curve.

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 shapes assumed by the yield curve. An extension offered by Svensson (1994) renders more flexibility. 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. Those reasons, combined with the empirical evidence for good in and out-of-sample forecasting performance promoted the acceptance of this model.

At the very heart of affine term structure models, lies the decomposition of the yield 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 might 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. They implement a bootstrap-based bias correction procedure and document that correcting the bias improves out-of-sample performance. Their approach focus on the bias of the factor persistence, not on bias in the intercept.

Usually, we target the yields themselves, factor modelling is a means to an end. Here I 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 can be applied to directly extract potential remains of information. As recently been suggested, conditional on the existence of such bias, this has the potential to improve forecasting performance.

I empirically illustrate this point using the NS model. The NS model is compared favourably in terms of forecasting performance to other less parsimonious models (e.g. Mönch, 2008). 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. One fairly standard way to do that is by adding a lagged residual to the model. While doing so might improve prediction, we lose the pleasant feature of factor interpretation, as the lagged residual will extract ex-

planatory power from the factors. At the same time, it is unattractive to squander potential information that can be used to improve forecast accuracy. Possible pragmatic and intuitive solution is to model the *out-of-sample* errors as an AR(1) process and forecast the error term. The resulting forecast is the original NS forecast with an added forecast of the error term. This extension preserves the attractive feature of the NS factors interpretation while significantly improves the quality of prediction throughout the curve. Taieb and Hyndman (2012) perform a similar two-stage procedure in a different context. A recursive multi-step ahead forecast is known to be biased even if the base model is correctly specified. They correct this bias in a similar fashion. They model the forecast errors and rectify the original biased forecast subsequently.

The rest of the paper is organised as follows. The next section motivates the adjustment, Section 3 presents the empirical results. Significance and robustness are discussed in section 4 where different term structure model specifications are considered. Section 5 concludes.

2 The model

We start with a general framework and notation, a general factor models, then move on to the more specific NS model which is used in the empirical exercise, ending the section with the outlined procedure for bias correction.

2.1 General factor model

A general dynamic factor model can be written as:

$$\mathbf{y}_t = \mathbf{B}(\boldsymbol{\theta})\mathbf{F}_t + \boldsymbol{\varepsilon}_t \quad (1)$$

$$\mathbf{F}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}(L)\mathbf{F}_{t-1} + \boldsymbol{\eta}_t \quad (2)$$

with \mathbf{y}_t vector of M yields at time t . $\mathbf{B}(\boldsymbol{\theta})$ is an $M \times r$ matrix of r factors loadings, possibly dependent on $\boldsymbol{\theta}$ parameters. $\boldsymbol{\Phi}(L)$ denotes $(r \times r) \times (r \times r)$ matrix of order p lag polynomials. $\boldsymbol{\mu}$ is $r \times 1$ vector of constants. Finally, $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are $M \times 1$ and $r \times 1$ vectors of residuals that are assumed to be uncorrelated. The time series aspect of the model in (2) describes the dynamics of the low dimensional vector of factors. It is common practice to use three factors for the yield curve (DL, 2006, Mönch, 2008, Tobias et al., 2012 for a recent counter example).

The cross sectional aspect, the mapping between factors and yields in (1), is quite flexible. We can reduce the number of free parameters by imposing structure on $\mathbf{B}(\boldsymbol{\theta})$. Examples are Duffie-Kan restrictions (Duffie and Kan, 1996), smoothing restrictions (Koopman and Van der Wel, 2010) and the NS model.

A standard assumption underlying this class of factor models is the orthogonality of \mathbf{F}_t and $\boldsymbol{\varepsilon}_t$. This means that the yield curve is decomposed into a common component and an idiosyncratic component. When the factors are extracted using principal component method the loadings can be determined using least squares regression in (1). Another option is to estimate the model in a state-space framework using the Kalman Filter algorithm, maximizing the likelihood function with respect to the parameters. A recent example in our context is Koopman et al. (2010). This approach is more efficient under correct specification of the maximum likelihood function. However, the likelihood surface is highly non-linear with many local optima which complicates this undertaking. Numerical search can be sensitive to different search algorithms as well as starting values. In a simple simulation exercise, Hamilton and Wu (2012) illustrate the difficulty. Out of 100 different starting values provided, only one established global MLE, 18 did not progress from the initial starting values provided and the remaining 81 that satisfied convergence criterion ended up at a large range of local optima. In addition, Yu and Zivot (2010) find that using the state-space estimation method results in poor out-of-sample performance compared with the two-step procedure. Thus, the projection approach suggested by DL (2006) seems like a reasonable choice.

2.2 Dynamic Nelson-Siegel model

For the yield curve of interest rates, using the well known latent factor model suggested by Nelson and Siegel (1987), the loadings are predetermined functions of maturity τ . The representation given by DL (2006) to this model is given by:

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

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 "level" 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. In the special case where $\lambda_t = \lambda \forall t$, the factors β_t are obtained using a simple cross sectional regression across available maturities at time t . The residuals $\epsilon_t = \{\epsilon_{t,1}, \dots, \epsilon_{t,M}\}$ are assumed independent, to have mean zero and constant variance σ_t^2 at any time point t . Note that these assumptions 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.

The model is widely used by central banks (BIS 2005), and is a natural candidate for forecasting since the factors are strongly correlated over time. A strategy which has been shown empirically to perform very well, is to extract the factors using a cross sectional regression, forecast the future factors using (2) and use the relation in (1) to obtain the yield curve forecast. The factors β_t are interpreted as the level, slope and curvature with the corresponding loadings.

The h -step-ahead prediction is given by:

$$\hat{y}_{t+h}(\tau) = \hat{\beta}_{1,t+h} + \hat{\beta}_{2,t+h} \left(\frac{1 - \exp(-\lambda_t \tau)}{\lambda_t \tau} \right) + \hat{\beta}_{3,t+h} \left(\frac{1 - \exp(-\lambda_t \tau)}{\lambda_t \tau} - \exp(-\lambda_t \tau) \right) \quad (4)$$

with

$$\hat{\beta}_{t+h} = \hat{\alpha} + \hat{\Gamma} \hat{\beta}_t, \quad (5)$$

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

More detailed description and factor intuition can be found in DL 2006 and Koopman et al.,2010.

2.3 Bias correction procedure

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

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

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:

$$\hat{e}_{t+h}(\tau) = \delta(\tau, h) + \rho(\tau, h)e_t(\tau) \quad (6)$$

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:

$$\hat{y}_{t+h}^{adj}(\tau) = \hat{y}_{t+h}(\tau) + \hat{e}_{t+h}(\tau) = \hat{y}_{t+h}(\tau) + \hat{\delta}(\tau, h) + \hat{\rho}(\tau, h)\hat{e}_t(\tau) \quad (7)$$

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. Note that "model errors" refers to the out-of-sample errors, to stress the fact the coefficients are trained towards their direct use in forecasting. This is useful since the prediction bias can assume different structure for different forecasting horizons, which is directly accounted for.

A wrinkle in the proposal is that when there is no information to be extracted, i.e. the errors are in fact white noise, the extension is redundant and will result in a more noisy forecast caused by the added estimation uncertainty. In general, unnecessary additional estimation uncertainty diminishes forecast accuracy. In the next section empirical results

alleviate this concern.

3 Empirical results

In this section I 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.. The last data point in their dataset is 12/2000. The data is therefore complement with subsequent months until 12/2009. The additional 108 monthly observations are taken from CRSP unsmoothed Fama and Bliss (1987) forward rates and were converted into unsmoothed Fama-Bliss zero yields in the same way as in DL (2006). Same data was also used in Jungbacker et al. (2012) and Koopman and van der wel (2011). Table 3.1 presents selected summary statistics of the data.

The data can be downloaded from <http://www.ssc.upenn.edu/~fdiebold/papers/paper49/FBFITTED.txt>

Maturity	Mean	Sd	Skewness	Kurtosis	Min	Max	$\hat{\rho}(1)$	$\hat{\rho}(12)$
3	4.465	2.208	-0.224	2.349	0.041	9.131	0.982	0.644
6	4.610	2.211	-0.225	2.343	0.150	9.324	0.981	0.643
9	4.717	2.217	-0.202	2.339	0.193	9.343	0.980	0.648
12	4.840	2.250	-0.187	2.326	0.245	9.683	0.979	0.656
15	4.957	2.285	-0.176	2.322	0.377	9.988	0.979	0.662
18	5.036	2.276	-0.150	2.343	0.438	10.188	0.979	0.664
21	5.107	2.256	-0.126	2.349	0.532	10.274	0.978	0.666
24	5.146	2.223	-0.103	2.371	0.532	10.413	0.977	0.664
30	5.290	2.214	-0.055	2.400	0.819	10.748	0.976	0.670
36	5.401	2.172	-0.014	2.408	0.978	10.787	0.975	0.673
48	5.615	2.130	0.083	2.484	1.019	11.269	0.974	0.676
60	5.750	2.058	0.186	2.464	1.556	11.313	0.973	0.682
72	5.905	2.060	0.272	2.571	1.525	11.653	0.973	0.682
84	6.011	1.985	0.343	2.577	2.179	11.841	0.971	0.684
96	6.110	1.954	0.334	2.462	2.105	11.512	0.973	0.696
108	6.172	1.936	0.376	2.482	2.152	11.664	0.973	0.698
120	6.192	1.892	0.442	2.494	2.679	11.663	0.972	0.694

Table 1: The table reports summary statistics for U.S. treasury yields over the period 1985:01 to 2009:12. Monthly frequency data, constructed using the unsmoothed Fama-Bliss method. For each maturity (measured in months) we observe the mean, standard deviation (Sd), Minimum (Min), Maximum (Max) and sample autocorrelation for one and twelve lags ($\hat{\rho}(\cdot)$).

3.2 Residuals from a Nelson-Siegel model

In order to make the case for the correction procedure, we examine the in-sample residuals Nelson-Siegel fit, as in DL (2006). Table 2 presents the descriptives statistics and shows that there is considerable autocorrelation for all maturities. This might be a result from illiquidity in the traded bonds. Such persistence in the residuals is also documented by de Jong (2000) and Bliss (1997b). In addition, we can see that the mean, which under the factor modelling framework is assumed to be zero, is higher for some maturities while lower for others. This has been tested formally using the standard *t-test*. The correction proposed in (6) uses δ to capture mean deviation and ρ to capture error persistence. The decomposition of the yield curve into its common part and residual part implies that null forecasting power remains in the residuals. In such a case the departure from the white noise assumption observed in table 2, namely the rejection of the mean zero assumption and the high autocorrelation,

Maturity	Mean	Sd	Min	Max	$\hat{\rho}(1)$	$\hat{\rho}(12)$
3	-0.018	0.080	-0.332	0.156	0.776	0.157
6	-0.013	0.042	-0.141	0.218	0.290	0.257
9	-0.026	0.062	-0.201	0.218	0.699	0.218
12	0.013	0.080	-0.160	0.267	0.568	0.323
15	0.063	0.050	-0.063	0.242	0.657	0.139
18	0.048	0.035	-0.048	0.165	0.495	0.183
21	0.026	0.030	-0.091	0.126	0.356	-0.062
24	-0.028	0.045	-0.190	0.082	0.657	0.214
30	-0.017	0.036	-0.200	0.098	0.377	0.072
36	-0.037	0.046	-0.203	0.128	0.598	0.052
48	-0.019	0.065	-0.204	0.230	0.731	0.229
60	-0.052	0.058	-0.199	0.186	0.756	-0.013
72	0.010	0.080	-0.134	0.399	0.903	0.289
84	0.003	0.066	-0.259	0.337	0.616	-0.033
96	0.032	0.045	-0.202	0.111	0.727	0.165
108	0.033	0.046	-0.161	0.132	0.667	0.072
120	-0.017	0.072	-0.256	0.164	0.625	0.232

Table 2: The table reports summary statistics for yield curve residuals from the Nelson Siegel model fit for the period 1985:1 to 2000:12. For each maturity (measured in months) we observe mean, standard deviation (Sd), Minimum (Min), Maximum (Max) and sample autocorrelation for one and twelve lags ($\hat{\rho}(\cdot)$). Bold numbers represent a rejection of the residual mean being equal to zero, at 5% level using a *t-test*. Ljung Box test for the NULL of no autocorrelation is firmly rejected for all yields.

can be reconciled by the relatively short time span considered. Yields are very persistent so the data contains only a handful of business cycles. Furthermore, incorporating residuals into a forecasting model should not result in any forecast accuracy gains. In contrast, if the underlying model assumptions about the nature of the residuals being white noise is only an approximation or altogether false, we might be able to extract remaining prediction power from the errors to improve our out-of-sample performance, while at the same time preserving the natural factor model structure and interpretation. Table 2 is useful to communicate the motivation for the adjustment. Note that I do not model the in-sample residuals, but the out-of-sample forecasting error. I train the estimators towards their direct target and so for example, if the bias in mean of the forecast errors for long forecasting horizon is larger than that of short forecasting horizon, this will be reflected in the corresponding estimates.

3.3 Estimation

In the estimation procedure, I follow closely DL (2006), so as to provide consistent comparison and in order to avoid any ad-hoc choices which are not properly backed by existing economic theory, e.g. rolling versus expanding window. The in-sample period starts at 01/1985 and the first forecast is made for 01/1994. Parameters are estimated recursively with additional observation added each month. Factors are modelled as an AR(1) process, i.e. $\hat{\mathbf{F}}$ in (5) is restricted to be diagonal, λ_t is not treated as time varying but is set to $\lambda = 0.0609$, the value that maximizes the third factor loading at maturity of 30 months. In a subsequent section, we relax these constraints and examine other specifications.

First we create the forecasts for the yields using the NS model as described above. We use the first 30 forecast errors (prior to 01/1994), to obtain initial estimates for δ and ρ parameters in (6), after which estimation proceeds in a recursive manner. Once the forecast for the out-of-sample error is made, the original NS forecast is adjusted according to (7).

3.4 Out-of-sample forecasting performance

The evaluation metric is root mean squared prediction error (RMSPE) as it is widely used in forecasting practice. Table 3 presents the results from the pseudo out-of-sample forecast exercise and compare the NS model with the Bias Corrected NS (BCNS).

Maturity	One month ahead		Six month ahead		Twelve month ahead	
	NS	BCNS	NS	BCNS	NS	BCNS
3 Months	0.283	0.242	1.002	0.776	1.642	1.501
6 Months	0.251	0.223	1.005	0.806	1.629	1.472
12 Months	0.269	0.246	1.003	0.848	1.594	1.396
36 Months	0.316	0.296	0.950	0.850	1.440	1.147
60 Months	0.315	0.295	0.879	0.782	1.316	0.981
120 Months	0.289	0.287	0.713	0.636	1.120	0.777

Table 3: Out-of-sample root mean squared prediction error (in percentage points) without (NS) and with (BCNS) our proposed adjustment. The forecast period is 1994:01 - 2009:12, for one, six and twelve month ahead forecasts. NS stands for the Nelson-Siegel model. BCNS stands for bias-corrected Nelson-Siegel model.

We observe that by utilizing the proposed correction, we achieve better results throughout the curve. This is to be expected from analysis presented in the earlier section and in line with

recent literature. The improvement in performance stems from the fact that the NS model has substantial forecast errors autocorrelation and mean which is different from zero, these facts are utilized in the proposed adjustment. The improvement suggests that the bias is strong enough to compensate for added the estimation noise from the extra parameters δ and ρ . Also, note that the adjustment *uniformly* does not impair accuracy which is comforting in the minmax sense. When the adjustment does not improve out-of-sample performance, on average, performance is not diminished either. It is a further indication that the proposed adjustment is indeed agreeable.

4 Significance testing and robustness checks

This section examines whether improvement presented in the earlier section is large. In addition, we perform the analysis for different time periods and investigate the effectiveness of the proposed adjustment using different model specifications. Specifically, (1) we allow for conditional correlated factors (CFNS), (2) we allow for time varying loadings (TVL) using a cross sectional non-linear least squares as was proposed by DL, and (3) we add a second hump term to the NS model resulting in the time varying loading Neslon-Siegel-Svensson model (NSS).

Allowing for cross sectional relation between factors was done by Mönch (2008) and is added here for completeness. Time varying loadings specification was recently investigated in Koopman et al. (2010), although not in the context of forecasting which is of particular interest here. The time varying loading Neslon-Siegel-Svensson extension is brought here for the first time. The goal is to illustrate that the proposal put forward here is effective regardless of model specification. Notwithstanding, an examination of the forecasting performance from the latter two specifications, which are more flexible and complex, may be interesting to look at in their own right.

4.1 Clark-West test

The proposed extension is nested in original model, i.e. when $\delta = \rho = 0$ the model boils down to the original one. In case these parameter are redundant, the adjustment is not needed and essentially adds estimation noise to the model. A natural result of such added

estimation noise is worse out-of-sample forecasting performance. While our results are actually an improvement in out-of-sample performance, we now check whether this improvement is indeed large. To this end we use the Clark-West test (2007) for nested models. Results are statistically significant for all three forecasting horizon considered. This result supports the hypothesis that for prediction purposes, the correction procedure proposed here is a useful addition which leads to superior forecasting performance.

Maturity	One month ahead	Six month ahead	Twelve months ahead
3 Months	3.75	3.14	2.02
6 Months	4.18	3.32	2.00
12 Months	3.82	3.22	2.10
36 Months	3.72	3.28	3.19
60 Months	4.15	3.59	4.29
120 Months	2.13	3.47	5.31

Table 4: Clark-West test statistics for selected maturities for the three forecasting horizons considered. It is a one-sided test with critical value of 1.65 for 5% significance level. Test-statistics for forecasting horizon longer than one were computed using the Newey-West correction.

4.2 Different subperiods analysis

In order to verify that the results are not driven by any particular period, I divide the dataset into three non over-lapping subperiods and carry out the same experiment. Table 5 summarizes the results which match previous argumentation.

In some periods forecasting gains are higher than others. That said, there is little to no evidence of worse performance due to estimation of redundant parameters, as is expected in the case where the errors are, in fact, white noise processes. The results are as expected from the significance testing performed in the previous subsection.

4.3 Different model specifications

Analysis presented previously followed specifications as they appear in DL (2006). I turn to examine the proposed correction under different specifications. This is done to confirm that

We also performed the Giacomini-White test (2006) which does not presume under the NULL worse out-of-sample performance of the nesting model, but merely compares the difference using a *t-test* approach. Results tell the same story.

Forecasting Performance Subperiod 1994 - 1998

Maturity	One month ahead		Six month ahead		Twelve month ahead	
	NS	BCNS	NS	BCNS	NS	BCNS
3 Months	0.174	0.159	0.545	0.601	0.861	1.061
6 Months	0.200	0.174	0.617	0.686	0.834	0.970
12 Months	0.244	0.222	0.684	0.801	0.849	0.918
36 Months	0.291	0.272	0.824	0.845	1.118	0.899
60 Months	0.305	0.273	0.880	0.790	1.277	0.871
120 Months	0.268	0.259	0.846	0.707	1.355	0.820

Forecasting Performance Subperiod 1999 - 2003

Maturity	One month ahead		Six month ahead		Twelve month ahead	
	NS	BCNS	NS	BCNS	NS	BCNS
3 Months	0.247	0.216	1.163	0.865	1.964	1.635
6 Months	0.273	0.224	1.211	0.861	2.015	1.651
12 Months	0.298	0.252	1.222	0.847	2.021	1.599
36 Months	0.324	0.315	1.077	0.871	1.738	1.320
60 Months	0.324	0.318	0.965	0.832	1.509	1.134
120 Months	0.306	0.293	0.763	0.662	1.206	0.846

Forecasting Performance Subperiod 2004 - 2009

Maturity	One month ahead		Six month ahead		Twelve month ahead	
	NS	BCNS	NS	BCNS	NS	BCNS
3 Months	0.367	0.306	1.106	0.801	1.775	1.607
6 Months	0.266	0.252	1.040	0.830	1.681	1.564
12 Months	0.260	0.257	0.993	0.875	1.568	1.455
36 Months	0.327	0.295	0.919	0.831	1.314	1.123
60 Months	0.313	0.289	0.795	0.726	1.088	0.899
120 Months	0.286	0.299	0.546	0.555	0.738	0.682

Table 5: Out-of-sample root mean squared prediction error (in percentage points) without (NS) and with (BCNS) our proposed adjustment. The three sub-periods: 1994:01 - 1998:12, 1999:01 - 2003:12, 2004:01 - 2009:12. One, six and twelve month ahead forecasts.

our proposal is not performing well only under a specific configuration but breaks down when we allow more flexibility in the factor modelling stage. I conduct the empirical experiment using different additional adaptations. First, relax the diagonal restriction on $\hat{\Gamma}$ in (5) to allow conditional cross-factor interaction. Differently put, factors follow a VAR process as oppose to a univariate AR's. Second, allow λ to vary over time. The parameter λ sets the shape of the curve. In related literature, it is customary to fix λ at some constant throughout

the analysis. In the analysis performed earlier in the paper λ was set to 0.0609 as in DL (2006). Other examples for fixing λ include Diebold, Rudebusch, and Aruoba (2006), Mönch (2008) and more recently, Van Dijk et al. (2012). In 2010, Koopman et al. find evidence for time variation in λ . They propose to integrate λ as a fourth factor into the VAR process in (5). This will allow for a more flexible dynamics in the factor dependency structure. In the same vein, we can achieve even more flexibility by adding a fourth term in (3), a second hump-shape (U shape). This will undoubtedly improve in-sample fit and perhaps provide better out-of-sample results. I assess these suggestions in the context of forecasting, which is an analysis of independent interest.

4.3.1 Correlated factors NS (CFNS)

The difference between the NS and the CFNS is that for the CFNS, the Γ matrix in (5) is fully parametrized, as oppose to diagonal as was advocated by DL (2006). This allows possible conditional interaction between factors. Put another way, factors follow a VAR(1) process instead of an three AR(1) processes. As can be seen from Table 6, the reduction in RMSPE from the bias correction step outlined is maintained. Note that at times where the CFNS model is more accurate, the gain from the bias correction step is not as large compared with the original NS (using AR dynamics for the factors). For example, for a one moth ahead prediction of the yield with three months to maturity, the CFNS performs much better than the NS. In this case, the gain from out-of-sample error forecasting, and the adjustment stage is not as large as for the original NS specification.

Maturity	One month ahead		Six month ahead		Twelve month ahead	
	CFNS	BCCFNS	CFNS	BCCFNS	CFNS	BCCFNS
3 Months	0.245	0.226	0.872	0.777	1.642	1.465
6 Months	0.222	0.212	0.924	0.837	1.701	1.538
12 Months	0.255	0.246	0.98	0.905	1.75	1.57
36 Months	0.314	0.299	1.014	0.897	1.715	1.41
60 Months	0.316	0.297	0.959	0.831	1.587	1.237
120 Months	0.292	0.289	0.79	0.692	1.341	1.008

Table 6: Out-of-sample root mean squared prediction error (in percentage points) without (CFNS) and with (BCCFNS) our proposed bias-correction procedure. The forecast period is 1994:01 - 2009:12, for one, six and twelve month ahead forecasts. CFNS stands for correlated factors NS model. BCCFNS stands for bias-corrected correlated NS model.

4.3.2 Time varying loadings (TVL)

Given λ , factor estimation is a simple cross sectional least-squares as described in (3). However, in 2010, Koopman et al. find evidence for time variation in λ . This can be accounted for, as was proposed by DL (2006). The cross sectional factor estimation procedure becomes non-linear least squares. The estimation is slightly complicated by the fact that at each point in time we have relatively few observations. Too few degrees of freedom can cause instability in estimation. To show how to mitigate the problem, denote:

$$\mathbf{X}_t = \begin{pmatrix} 1 & \left(\frac{1-e^{-\lambda_t\tau_1}}{\lambda_t\tau_1}\right) & \left(\frac{1-e^{-\lambda_t\tau_1}}{\lambda_t\tau_1} - e^{-\lambda_t\tau_1}\right) \\ \vdots & \vdots & \vdots \\ 1 & \left(\frac{1-e^{-\lambda_t\tau_M}}{\lambda_t\tau_M}\right) & \left(\frac{1-e^{-\lambda_t\tau_M}}{\lambda_t\tau_M} - e^{-\lambda_t\tau_M}\right) \end{pmatrix}, \quad \text{and} \quad \mathbf{y}_t = \begin{pmatrix} y_{t,1}(\tau_1) \\ \vdots \\ y_{t,M}(\tau_M) \end{pmatrix}.$$

Given λ_t , the best fit for the curve in the mean squared error sense is given by the three factors

$$\boldsymbol{\beta}_t(\lambda_t) = (\mathbf{X}_t' \mathbf{X}_t)^{-1} \mathbf{X}_t' \mathbf{y}_t,$$

In each period, I grid search over λ_t for the vector $\boldsymbol{\beta}_t(\lambda_t)$ that minimizes the sum of squared errors in (3). The result from the coarse grid search is then provided as starting value for the search algorithm.

The four factors:

$$\boldsymbol{\theta}_t = \{\beta_{t,1}, \beta_{t,2}, \beta_{t,3}, \lambda_t\} \tag{8}$$

follow an unrestricted VAR,

$$\boldsymbol{\theta}_t = \boldsymbol{\kappa} + \boldsymbol{\Delta} \boldsymbol{\theta}_{t-1} + \mathbf{v}_t. \tag{9}$$

The dimensions of $\boldsymbol{\kappa}$, $\boldsymbol{\Delta}$, and \mathbf{v}_t are 4×1 , 4×4 , and 4×1 , respectively. As before, the estimation is performed recursively with each additional time point. The yield forecast is given by its mapping from (4) using the additional forecast term $\hat{\lambda}_t$ together with $\hat{\boldsymbol{\beta}}_t$.

Results shown in table 7 accord with the pattern previously observed. The bias correction improves accuracy. At times less, but overall across the board. Importantly, accuracy never decreases as a result of the proposed adjustment.

I use "Brent" algorithm in the open source *R* system for statistical computing

Maturity	One month ahead		Six month ahead		Twelve month ahead	
	TVL	BCTVL	TVL	BCTVL	TVL	BCTVL
3 Months	0.324	0.280	0.976	0.870	1.742	1.624
6 Months	0.249	0.232	0.996	0.894	1.781	1.623
12 Months	0.287	0.258	1.038	0.940	1.806	1.593
36 Months	0.361	0.312	1.024	0.923	1.670	1.394
60 Months	0.344	0.325	0.938	0.865	1.483	1.221
120 Months	0.438	0.398	0.814	0.799	1.251	1.062

Table 7: Out-of-sample root mean squared prediction error (in percentage points) without (TVL) and with (BCTVL) our proposed adjustment. The forecast period is 1994:01 - 2009:12, for one, six and twelve month ahead forecasts. TVL stands for time varying loadings. BCTVL stands for bias-corrected time varying loadings.

4.3.3 Dynamic Nelson-Siegel-Svensson model (NSS)

On average, we expect a simple relation between yields and maturities. Namely, the longer the buyer of the bond is willing to wait, the higher the compensation. When this is the case, the Nelson-Siegel model is adequate and provides satisfactory results. However, in some periods, data show more complex yield-maturity relation and the curve assumes irregular shapes. When those occur, a more flexible model is needed. Svensson (1994) suggested adding an extra term to improve flexibility. The cross sectional mapping is now given by:

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t} \left(\frac{1 - \exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau} \right) + \beta_{3,t} \left(\frac{1 - \exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau} - \exp(-\lambda_{1,t}\tau) \right) + \beta_{4,t} \left(\frac{1 - \exp(-\lambda_{2,t}\tau)}{\lambda_{2,t}\tau} - \exp(-\lambda_{2,t}\tau) \right) + \varepsilon_t. \quad (10)$$

Svensson (1994) shows that this additional hump-shape factor improves the fit of the curve. Analogously to the parameters β_3 and λ_1 , the parameters β_4 and λ_2 determine the size and location of the second hump respectively. Calibration of the model was done in a similar fashion to the TVL model described above. Dynamics in the time series aspect is now more complex. The parameter vector $\boldsymbol{\theta}_t$ in (8) now contains two extra terms, $\beta_{t,4}$ and $\lambda_{t,2}$. Hence a system of six equations as oppose to the three we have in the NS model and four in the time varying loadings case. When we allow cross relation between the six components in $\boldsymbol{\theta}$, we need to estimate $36 + 6$ parameters for factors point forecasts, which will substantially

increase estimation uncertainty. This model balances between the number of periods for which the extra term is beneficial and the added estimation noise which potentially decrease forecast accuracy. If only few periods benefit from this extra hump-shape factor, we can expect an overall worse out-of-sample results compared with the more parsimonious NS model.

Maturity	One month ahead		Six month ahead		Twelve month ahead	
	NSS	BCNSS	NSS	BCNSS	NSS	BCNSS
3 Months	0.352	0.318	0.976	0.87	1.665	1.68
6 Months	0.394	0.309	1.057	0.941	1.71	1.683
12 Months	0.414	0.334	1.103	0.989	1.726	1.618
36 Months	0.401	0.368	1.004	0.926	1.522	1.241
60 Months	0.443	0.392	0.918	0.891	1.326	1.037
120 Months	0.503	0.412	0.843	0.829	1.081	0.897

Table 8: The table presents the out-of-sample root mean squared prediction error (in percentage points) without (NSS) and with (BCNSS) our proposed adjustment. The forecast period is 1994:01 - 2009:12, for one, six and twelve month ahead forecasts. The model is estimated using time varying loadings. NSS stands for Nelson-Siegel-Svensson model. BCNSS stands for bias-corrected Nelson-Siegel-Svensson model.

The table 8 bears resemblance to previous results presented, which reinforces the main point of the paper. Interestingly, the added flexibility, which was shown to provide much better in-sample fit by Svensson (1994), produces much worse out-of-sample performance for the short horizon and slightly worse for longer forecasting horizons, compared with the simple DL (2006) specification.

Overall, results described in this section support the line of reasoning presented earlier. The correction step generally improves, and rarely harms prediction accuracy. We also observe the more flexible model configurations, such as the time varying loadings and the dynamic NSS model which provides better fit in-sample, performs badly out-of-sample compared with the more parsimonious alternatives. Yu and Zivot (2010) corroborate this result using corporate bonds data. A conclusion from this analysis is that allowing for time varying loadings, in terms of forecasting, can be labelled as excessive.

We also performed the exercise using AR dynamics, results are less competitive and so are omitted.

5 Concluding remarks

I propose a way to improve out of sample forecasting performance for the yield curve. With a focus set on out-of-sample performance I show that when the yield curve is modelled using a factor model, the residuals may still contain relevant information. This information can be exploited in favour of a more accurate prediction. That is, the factor model can be augmented with an additional step to extract remains of information from the model errors. I empirically illustrate this point using the NS model of DL (2006). Large improvement in forecasting performance is achieved across the curve and for different forecasting horizons. Results are significant and robust to three sub-samples for different time periods as well as for different model specifications. Although exemplified here using the NS model, the approach is general and can be applied whenever a factor model is used for yields modelling and forecasting. The correction step does not vary with model choice, thus for any specific model configuration, the procedure can be back-tested for possible future implementation. The bias correction step does not involve time-consuming optimization scheme, thereby making it well suited for practical applications.

Acknowledgements

The author thank Kees E. Bouwman and Dick van Dijk for commenting on several previous versions of the paper. In addition, the participants of EIPC 2012 conference in Rotterdam, in particular, Michel van der Wel and Dennis Karstanje for helpful discussions.

References

- Abadir, K., & Magnus, J. (2002). Notation in econometrics: a proposal for a standard. *The Econometrics Journal*, 5(1), 76–90.
- Adrian, T., Crump, R., & Moench, E. (2012). Pricing the term structure with linear regressions. *FRB of New York Staff Report*(340).
- Ang, A., & Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics*, 50(4), 745–787.
- Bauer, M., Rudebusch, G., & Wu, J. (2012). Correcting estimation bias in dynamic term structure models. *Journal of Business & Economic Statistics*, 30(3), 454–467.
- Bliss, R. (1997). Movements in the term structure of interest rates. *Economic Review*(Q 4), 16–33.
- Christensen, J., Diebold, F., & Rudebusch, G. (2008). *An arbitrage-free generalized nelson-siegel term structure model* (Tech. Rep.). National Bureau of Economic Research.
- Clark, T., & West, K. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1), 291–311.
- Coroneo, L., Nyholm, K., & Vidova-Koleva, R. (2011). How arbitrage-free is the nelson-siegel model? *Journal of Empirical Finance*, 18(3), 393–407.
- Cox, J., Ingersoll Jr, J., & Ross, S. (1985). An intertemporal general equilibrium model of asset prices. *Econometrica: Journal of the Econometric Society*, 363–384.
- De Jong, F. (2000). Time series and cross-section information in affine term-structure models. *Journal of Business & Economic Statistics*, 300–314.
- De Pooter, M., Ravazzolo, F., & Van Dijk, D. (2010). Term structure forecasting using macro factors and forecast combination. *FRB International Finance Discussion Paper*(993).
- Diebold, F., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2), 337–364.
- Diebold, F., Rudebusch, G., & Borağan Aruoba, S. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of econometrics*, 131(1), 309–338.
- Duffee, G. (2011). *Forecasting with the term structure: The role of no-arbitrage restric-*

- tions (Tech. Rep.). Working papers//the Johns Hopkins University, Department of Economics.
- Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance*, 57, 405-443.
- Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical finance*, 6(4), 379-406.
- Exterkate, P., Dijk, D., Heij, C., & Groenen, P. (2011). Forecasting the yield curve in a data-rich environment using the factor-augmented nelson-siegel model. *Journal of Forecasting*.
- Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545-1578.
- Gilli, M., Große, S., & Schumann, E. (2010). Calibrating the nelson-siegel-svensson model. *Available at SSRN 1676747*.
- Hamilton, J., & Wu, J. (2011). *Testable implications of affine term structure models* (Tech. Rep.). National Bureau of Economic Research.
- Hamilton, J., & Wu, J. (2012). Identification and estimation of gaussian affine term structure models. *Journal of Econometrics*.
- Heath, D., Jarrow, R., & Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society*, 77-105.
- Hull, J., & White, A. (1990). Pricing interest-rate-derivative securities. *Review of financial studies*, 3(4), 573-592.
- Jungbacker, K. S. J., Borus, & Wel, M. Van der. (2012). Dynamic factor analysis with application to the u.s. term structure of interest rates. *Journal of Applied Econometrics*, *Forthcoming*.
- Koopman, S., Mallee, M., & Wel, M. Van der. (2010). Analyzing the term structure of interest rates using the dynamic nelson-siegel model with time-varying parameters. *Journal of Business & Economic Statistics*, 28(3), 329-343.
- Koopman, S., & Wel, M. Van der. (2011). Forecasting the us term structure of interest rates using a macroeconomic smooth dynamic factor model. *Available at SSRN 1804930*.

- Litterman, R., & Scheinkman, J. (1991). Common factors affecting bond returns. *The Journal of Fixed Income*, 1(1), 54–61.
- Mönch, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented var approach. *Journal of Econometrics*, 146(1), 26–43.
- Nelson, C., & Siegel, A. (1988). *Parsimonious modeling of yield curves for us treasury bills*. National Bureau of Economic Research Cambridge, Mass., USA.
- Settlements, B. f. I. (2005). *Zero-coupon yield curves: technical documentation*. Bank for International Settlements. Available from <http://EconPapers.repec.org/RePEc:bis:bisbps:25>
- Svensson, L. (1994). *Estimating and interpreting forward interest rates: Sweden 1992-1994* (Tech. Rep.). National Bureau of Economic Research.
- Taieb, S., & Hyndman, R. (2012). Recursive and direct multi-step forecasting: the best of both worlds. *Working paper*.
- Van Dijk, D., Koopman, S., Wel, M. Van der, & Wright, J. (2012). Forecasting interest rates with shifting endpoints.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of financial economics*, 5(2), 177–188.
- Yu, W.-C., & Salyards, D. M. (2009). Parsimonious modeling and forecasting of corporate yield curve. *Journal of Forecasting*, 28(1), 73-88. Available from <http://EconPapers.repec.org/RePEc:jof:jforec:v:28:y:2009:i:1:p:73-88>
- Yu, W.-C., & Zivot, E. W. (2010). Forecasting the term structures of treasury and corporate yields: Dynamic nelson-siegel models evaluation. *International Journal of Forecasting*, *Forthcoming*.