Prediction bias correction for dynamic term structure models

Eran Raviv

Econometric Institute, Erasmus University Rotterdam
Tinbergen Institute

February 4, 2015

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

*Eran Raviv, APG Asset management, Gustav Mahler Square 3, 1082 MS Amsterdam, The Netherlands. Email: eran.raviv@apg-am.nl.

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 computationally 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 favorably 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, $\boldsymbol{\tau} = \{\tau_1, ..., \tau_M\}$.

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 $\boldsymbol{\varepsilon}_t = \{\varepsilon_{t,1}, ..., \varepsilon_{t,M}\}$ 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 parametrised $\boldsymbol{\Gamma}$ 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

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

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:

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

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. 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 Diebold and Li (2006).

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 Diebold and Li (2006). Table 1 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 (1997). 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 (4) uses the parameter δ to capture mean deviation and the parameter 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. This is a clear departure from the white noise assumption commonly endorsed using this class of models.

[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.

Maturity	Mean	Sd	Min	Max	$\widehat{\rho}(1)$	$\widehat{\rho}(12)$
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 1: 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 ($\widehat{\rho}(\cdot)$). 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.

3.3 Estimation

In the estimation procedure, we follow closely Diebold and Li (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. $\widehat{\boldsymbol{\Gamma}}$ in (3) 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 an Appendix, we relax these constraints and show findings hold for 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 (4), 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 (5).

3.4 Out-of-sample forecasting performance

The evaluation metric is the widely used root mean squared prediction error (RMSPE). Table 2 presents the results from the out-of-sample forecast exercise and compare the NS model with the Bias Corrected NS (BCNS).For convenience, we present ratios of RMSPE to the RMSPE that is achieved using a simple random walk benchmark. A number smaller than one means that on average the model performs better than a random walk model.

Maturity	One month ahead		Six month ahead		Twelve month ahead	
	NS	BCNS	NS	BCNS	NS	BCNS
3 Months	1.165	0.996	1.116	0.864*	1.056	0.965
6 Months	1.055	0.937*	1.113	0.893*	1.060	0.958
12 Months	1.015	0.928*	1.102	0.932	1.083	0.948
36 Months	1.043	0.977	1.110	0.993	1.183	0.942
60 Months	1.050	0.983	1.117	0.994	1.252	0.933
120 Months	1.025	1.018	1.125	1.003	1.366	0.948

Table 2: Out-of-sample ratios of root mean squared prediction error without (NS) and with (BCNS) our proposed adjustment to the root mean squared prediction error from a simple random walk benchmark. 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 an earlier section. The improvement in performance stems from the 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 neither, indicating the additional step is indeed agreeable. Augmenting the model with the correction step brings significant improvement even over the random walk (RW) benchmark, a benchmark which is notoriously hard to beat (Altavilla et al., 2014). Entries with asterisks denote a significant improvement at the 5% significance level using the Diebold and Mariano (1995) test, accounting for autocorrelation using the HAC estimator for standard errors with truncation parameter set to forecasting horizon minus one. To make sure the results are not dominated by any specific time period, Figure 1 presents the 5 years rolling RMSPE with- and without the bias correction step for the twelve months forecasting horizon and for the different maturities. It is shown that accuracy gains in terms of RMSE are steady across time and along the curve.

4 Concluding remarks

With a focus set on out-of-sample performance we have shown that when the yield curve is modelled using a factor model, the residuals may still contain relevant information. This information is exploited towards a more accurate prediction. That is, the factor model is augmented with an additional step to extracts remains of information from the model errors. We empirically demonstrate this point using the widely used NS model of Diebold and Li (2006) . Large improvement in forecasting performance is achieved across the curve and for different forecasting horizons. In an Appendix it is also shown that results are significant and robust over time, as well as for different model specifications. One natural extension is going beyond point forecast, and examine whether the proposed adjustment also results in reduced forecast uncertainty. Although exemplified here using the NS model, the approach is general and can be applied whenever a factor model is used for

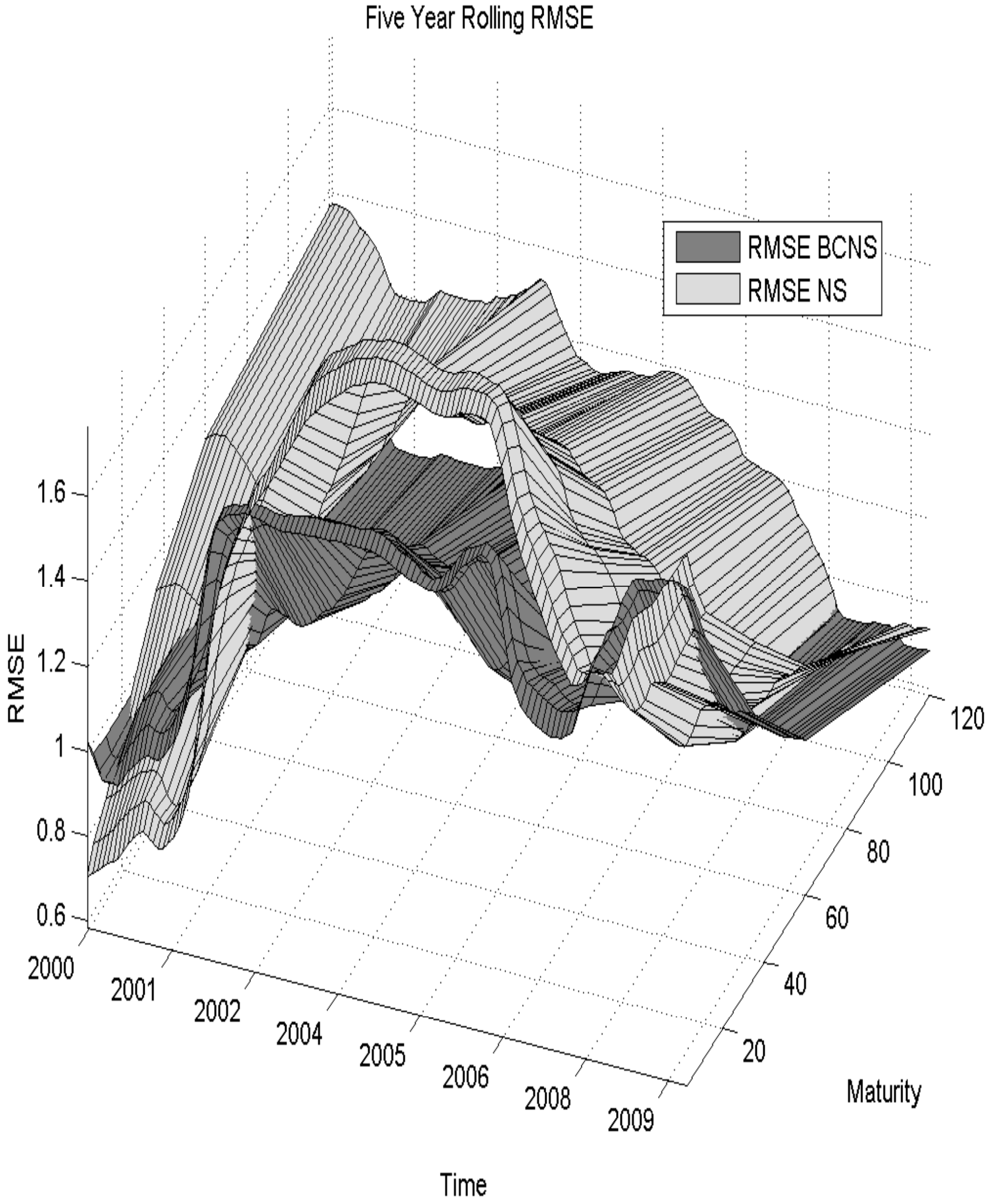

Figure 1: Five years rolling RMSPE for the different maturities, twelve month forecasting horizon, with our prosed bias correction step (RMSPE BCNS) and without (RMSPE NS).

yields modelling and forecasting. The correction step does not vary with model's choice and does not involve time-consuming optimization scheme, thereby making it especially well suited for practical applications that require out-of-sample accuracy via recursive estimation or rolling estimation schemes.

Acknowledgements

This paper has benefited from conversations with Dr. Kees E. Bouwman, Prof. Dick van Dijk, Dr. Michel van der Wel and Mr. Dennis Karstanje. I thank Dr. Pierre-Daniel Sarte and an anonymous referee for thoughtful and insightful comments, resulting in this final improved version.

References

Altavilla C, Giacomini R, Ragusa G. 2014. Anchoring the yield curve using survey expectations. *ECB Working Paper No. 1632.* .

Bauer M, Rudebusch G, Wu J. 2012. Correcting estimation bias in dynamic term structure models. *Journal of Business & Economic Statistics* **30**: 454–467.

Bliss R. 1997. Movements in the term structure of interest rates. *Economic Review* : 16–33.

Cox J, Ingersoll Jr J, Ross S. 1985. An intertemporal general equilibrium model of asset prices. *Econometrica: Journal of the Econometric Society* : 363–384.

Dai Q, Singleton KJ. 2002. Expectation puzzles, time-varying risk premia, and affine models of the term structure. *Journal of financial Economics* **63**: 415–441.

De Jong F. 2000. Time series and cross-section information in affine term-structure models. *Journal of Business & Economic Statistics* : 300–314.

Diebold F, Li C. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* **130**: 337–364.

Diebold F, Mariano R. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–265.

Diebold FX, Rudebusch GD. 2013. *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*. Princeton University Press.

Duffee G. 2011. Forecasting with the term structure: The role of no-arbitrage restrictions. Technical report, Working papers//the Johns Hopkins University, Department of Economics.

Duffie D, Kan R. 1996. A yield-factor model of interest rates. *Mathematical finance* **6**: 379–406.

Fama EF, Bliss RR. 1987. The information in long-maturity forward rates. *The American Economic Review* : 680–692.

Hamilton J, Wu J. 2011. Testable implications of affine term structure models. Technical report, National Bureau of Economic Research.

Hautsch N, Yang F. 2012. Bayesian inference in a stochastic volatility nelson–siegel model. *Computational Statistics & Data Analysis* **56**: 3774–3792.

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**: 573–592.

Litterman R, Scheinkman J. 1991. Common factors affecting bond returns. *The Journal of Fixed Income* **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**: 26–43.

Nelson C, Siegel A. 1988. Parsimoneous modeling of yield curves for us treasury bills.

Nelson CR, Siegel AF. 1987. Parsimonious modeling of yield curves. *Journal of business* : 473–489.

Vasicek O. 1977. An equilibrium characterization of the term structure. *Journal of financial economics* **5**: 177–188.

Appendix to "Prediction bias correction for dynamic term structure models"

Eran Raviv*

Econometric Institute, Erasmus University Rotterdam
Tinbergen Institute

February 4, 2015

Abstract

This Appendix includes additional information and results supplementing the paper. Section 1 describes the data used in the paper and presents additional motivation. Significance testing and different subperiods analysis are discussed in Section 2. Section 3 considers different model specification. Specifically, the Dynamic NS model using iterated rather the direct method applied in the paper, conditional correlated factors (CFNS), time varying loadings (TVL) using a cross sectional non-linear least squares as was proposed by DL, and we add a second hump term to the NS model resulting in the time varying loading Neslon-Siegel-Svensson model (NSS).

*Eran Raviv, APG Asset management, Gustav Mahler Square 3, 1082 MS Amsterdam, The Netherlands. Email: `eran.raviv@apg-am.nl`.

1 Data

We use the same data as in DL (2006), a balanced panel data of 17 maturities.[1]. 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 Diebold and Li (2006). Same data was also used in Jungbacker and Van der Wel (2012) and Koopman and Van der Wel (2011). Table 1 presents selected summary statistics of the data, and figure 1 presents the level, slope and curvature factors implied by the Nelson-Siegel model as specified in Diebold and Li (2006) along with the actual yields.

Maturity	Mean	Sd	Skewness	Kurtosis	Min	Max	$\widehat{\rho}(1)$	$\widehat{\rho}(12)$
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 ($\widehat{\rho}(\cdot)$).

2 Significance testing and subperiods analysis

This section examines whether improvement presented in the paper is indeed large, as well as presenting results for different time periods.

2.1 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

[1]The data can be downloaded from `http://www.ssc.upenn.edu/~fdiebold/papers/paper49/FBFITTED.txt`

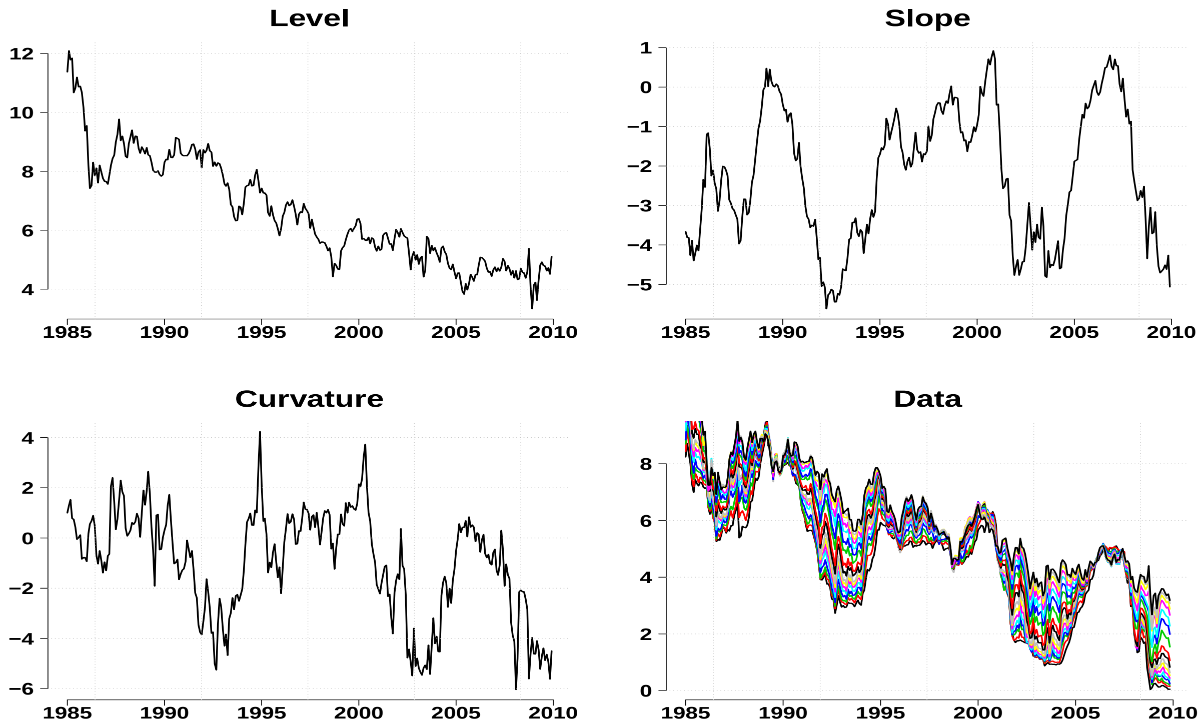

Figure 1: NS-based yield curve factors for the period from January 1985 to December 2009. **Top left**: level factor. **Top right**: slope factor. **Bottom left**: curvature factor. **Bottom right**: the underlying yields.

performance, we now check whether this improvement is indeed large. We use the Clark-West test (Clark and West, 2007) for nested models[2]. 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 2: 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.

[2]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.

2.2 Different subperiods analysis

In order to verify that the results are not driven by any particular period, we divide the dataset into three non over-lapping subperiods and carry out the same experiment. Table 3 summarizes the results which match previous argumentation.

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 3: 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.

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.

3 Different model specifications

We turn to examine the proposed correction under other model configuration. This is done to confirm that our proposal is not performing well only under a specific configuration but breaks down when we allow more flexibility in the factor modelling stage for example.

3.1 Direct vs. iterated factor forecasts

First, for a standard NS model it is not obvious that a direct method for forecasting the factors, i.e.:

$$\widehat{\boldsymbol{\beta}}_{t+h} = \widehat{\boldsymbol{\alpha}} + \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\beta}}_t, \tag{1}$$

should be the preferred method.[3] A viable and perhaps better alternative is to iterate the one-step horizon factor forecast forward to the desired forecasting horizon $h > 1$, i.e.:

$$\widehat{\boldsymbol{\beta}}_{t+h} = \widehat{\boldsymbol{\alpha}}_{(h=1)}(1 + \text{diag}\,(\sum_{j=1}^{h-1} \widehat{\boldsymbol{\Gamma}}^j_{(h=1)})) + \widehat{\boldsymbol{\Gamma}}^h_{(h=1)}\widehat{\boldsymbol{\beta}}_t, \tag{2}$$

where the diag$(\cdot)$ operator vectorises the diagonal, and we add the subscript $(\cdot)_{(h=1)}$ to distinguish the coefficients between the two equations 1 and 2. In general, if we assume martingale errors for the iterated approach, the vector $\widehat{\boldsymbol{\beta}}_{t+h}$ should converge to the same values as sample size increases in both approaches (Clark and McCracken, 2013). However, in small samples it is not clear which approach delivers better forecasting results. Once factor forecasts are attained, we proceed in the usual manner to obtain yields forecasts via:

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

Clearly, when $h = 1$ there is no difference between an iterated or a direct approach, therefore, and for ease of comparison table 4 presents the forecasting results for $h = 6, 12$ only:

Maturity	Six month ahead		Twelve month ahead	
	NS	BCNS	NS	BCNS
3 Months	0.962	0.775	0.987	0.925
6 Months	0.960	0.798	0.982	0.925
12 Months	0.964	0.834	0.990	0.919
36 Months	1.001	0.905	1.044	0.908
60 Months	1.011	0.912	1.078	0.890
120 Months	1.004	0.916	1.127	0.849

Table 4: Out-of-sample ratios of root mean squared prediction error without (NS) and with (BCNS) our proposed adjustment to the root mean squared prediction error from a simple random walk benchmark. The forecast period is 1994:01 - 2009:12, for six and twelve month ahead forecasts considering NS- forecasts based on the iterated approach. BCNS stands for bias-corrected Nelson-Siegel model. For the one moth ahead forecasts the columns are identical to those presented in the paper itself, and so are omitted here.

An interesting finding is that the iterated approach delivers better forecasting results than using the direct approach. This finding suggests that the iterated approach does not suffer much from accumulating bias in this case, while the direct approach, though applied more easily is less efficient in this case. We also observe that forecasting gains from the correction step remain.

We now conduct the empirical experiment using additional adaptations. First, relax the diagonal restriction on $\widehat{\Gamma}$ in the forecast equation 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

[3] I thank an anonymous referee for pointing this out.

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 Diebold and Li (2006). Other examples for fixing λ include Diebold et al. (2006), Mönch (2008) and van Dijk et al. (2014). In Koopman et al. (2010) find evidence for time variation in λ. They propose to integrate λ as a fourth factor into the VAR process. 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 to the original NS model, a second hump-shape (U shape). This will undoubtedly improve in-sample fit and perhaps provide better out-of-sample results. We assess these suggestions in the context of forecasting, which is of independent interest.

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. We aim 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.

3.2 Correlated factors NS (CFNS)

The difference between the NS and the CFNS is that for the CFNS, the Γ matrix in (3) in the original paper is fully parametrized, as oppose to diagonal as was advocated by Diebold and Li (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 5, the reduction in RMSPE from the outlined bias correction step is maintained. Note that at occasions 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 5: 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.

3.3 Time varying loadings (TVL)

Given λ, factor estimation is a simple cross sectional least-squares:

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t}\left(\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}\right) + \beta_{3,t}\left(\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)\right) + \varepsilon_t, \tag{4}$$

where available maturities at time t, $\boldsymbol{\tau} = \{\tau_1, ..., \tau_M\}$. However, in Koopman et al. (2010) find evidence for time variation in λ. This can be accounted for, as was proposed by Diebold and Li (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:

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

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

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

In each period, we grid search over λ_t for the vector $\boldsymbol{\beta}_t(\lambda_t)$ that minimizes the sum of squared errors in from the NS model. The result from the coarse grid search is then provided as starting value for the search algorithm[4].

The four factors:

$$\boldsymbol{\theta}_t = \{\beta_{t,1}, \beta_{t,2}, \beta_{t,3}, \lambda_t\} \tag{5}$$

follow an unrestricted VAR,

$$\boldsymbol{\theta}_t = \boldsymbol{\kappa} + \boldsymbol{\Delta}\boldsymbol{\theta}_{t-1} + \boldsymbol{\upsilon}_t. \tag{6}$$

The dimensions of $\boldsymbol{\kappa}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\upsilon}_t$ 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

$$\widehat{y}_{t+h}(\tau) = \widehat{\beta}_{1,t+h} + \widehat{\beta}_{2,t+h}\left(\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}\right) + \widehat{\beta}_{3,t+h}\left(\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)\right) \tag{7}$$

using the additional forecast term $\widehat{\lambda}_t$ together with $\widehat{\boldsymbol{\beta}}_t$.

Results shown in table 6 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.

3.4 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

[4]we 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 6: 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.

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:

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t}\left(\frac{1-\exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau}\right) + \beta_{3,t}\left(\frac{1-\exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau} - \exp(-\lambda_{1,t}\tau)\right) + \beta_{4,t}\left(\frac{1-\exp(-\lambda_{2,t}\tau)}{\lambda_{2,t}\tau} - \exp(-\lambda_{2,t}\tau)\right) + \varepsilon_t. \tag{8}$$

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

$$\boldsymbol{\theta}_t = \{\beta_{t,1}, \beta_{t,2}, \beta_{t,3}, \lambda_t\} \tag{9}$$

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.[5] 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.

The table 7 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 Diebold and Li (2006) specification.

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

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 7: 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 fro Nelson-Siegel-Svensson model. BCNSS stands for bias-corrected Nelson-Siegel-Svensson model.

References

Clark T, West K. 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* **138**: 291–311.

Clark TE, McCracken MW. 2013. Evaluating the accuracy of forecasts from vector autoregressions. *Advances in Econometrics* **32**: 117–168.

Diebold F, Li C. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* **130**: 337–364.

Diebold FX, Rudebusch GD, Boragan Aruoba S. 2006. The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of econometrics* **131**: 309–338.

Fama EF, Bliss RR. 1987. The information in long-maturity forward rates. *The American Economic Review* : 680–692.

Jungbacker KSJ Borus, Van der Wel M. 2012. Dynamic factor analysis with application to the u.s. term structure of interest rates. *Journal of Applied Econometrics* **Forthcoming**.

Koopman S, Mallee M, Van der Wel M. 2010. Analyzing the term structure of interest rates using the dynamic nelson–siegel model with time-varying parameters. *Journal of Business & Economic Statistics* **28**: 329–343.

Koopman S, Van der Wel M. 2011. Forecasting the us term structure of interest rates using a macroeconomic smooth dynamic factor model. *Available at SSRN 1804930* .

Mönch E. 2008. Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented var approach. *Journal of Econometrics* **146**: 26–43.

Svensson L. 1994. Estimating and interpreting forward interest rates: Sweden 1992-1994. Technical report, National Bureau of Economic Research.

van Dijk D, Koopman SJ, Wel M, Wright JH. 2014. Forecasting interest rates with shifting end-points. *Journal of Applied Econometrics* .