RESERVE BANK
OF NEW ZEALAND

DP2009/18

Forecasting New Zealand's economic growth using yield curve information

Leo Krippner and Leif Anders Thorsrud

December 2009

JEL classification: E43, E44, E47

www.rbnz.govt.nz/research/discusspapers/

Discussion Paper Series

ISSN 1177-7567

DP2009/18

Forecasting New Zealand's economic growth using yield curve information*

Leo Krippner and Leif Anders Thorsrud†

Abstract

We forecast economic growth in New Zealand using yield curve data within simple statistical models; i.e. typical OLS relationships that have been well-established for other countries, and related VAR specifications. We find that the yield curve data has significant forecasting power in absolute terms and performs well relative to various benchmarks. Specifications including measures of the yield curve slope produce the best forecasts overall. Our results also highlight the benefits of fully exploiting the timeliness of yield curve information (i.e it is always available and up to date).

* The Reserve Bank of New Zealand's discussion paper series is externally refereed. The views expressed in this paper are those of the author(s) and do not necessarily reflect the views of the Reserve Bank of New Zealand. The authors would like to thank Karsten Gerdrup, Francesco Ravazzolo, Christie Smith and members of the modeling team at the Reserve Bank of New Zealand for valuable comments and fruitful discussions, and one anonymous referee for providing comments on an earlier draft. John Groom provided welcome input on aspects of preparing the data. Any remaining errors are the authors's sole responsibility.

† Address: Economics Department, Reserve Bank of New Zealand, 2 The Terrace, PO Box 2498, Wellington, New Zealand. *Email addresses:* Leo.Krippner@rbnz.govt.nz and Leif-Anders.Thorsrud@rbnz.govt.nz

1 Introduction

This article investigates the use of yield curve data to forecast New Zealand's output growth. It is motivated by the ongoing development of the Reserve Bank's suite of statistical models for forecasting the New Zealand economy in real time, see for example Bloor (2009), and compelling evidence from the literature that relationships between the yield curve and future GDP growth are likely to prove useful in this regard.

Regarding the latter, the international literature linking the yield curve and GDP growth is far too extensive to cite in full,[1] so we simply highlight the relevant results for this article with selected examples. We also focus on the empirical strand of the literature, given our application is purely about forecasting, rather than taking a stand on any particular theoretical justification.[2]

Formal empirical research on the relationship between the slope of the yield curve (long-maturity less short-maturity rates) and economic growth began in the late 1980s. For example, Estrella and Hardouvelis (1991) established statistically significant single-equation relationships between the 10-year bond rate less 3-month bill rate, and lagged US economic growth or recession probabilities.[3] Since then, these empirical relationships have been regularly re-examined and confirmed for the US, and also established for other countries, notably Germany, Canada, and the United Kingdom; see, for example, Plosser and Rouwenhorst (1994), Estrella and Mishkin (1997) and Bernard and Gerlach (1998).

One path of more recent empirical literature has focussed on testing the out-of-sample forecasting performance of the yield curve. For example, using US data over the period 1875 to 1997 within single-equation rolling OLS regressions, Bordo and Haubrich (2008) find that the yield curve slope predicts future growth better

[1] A comprehensive bibliography by Estrella (2009) has more than 100 entries, and Wheelock and Wohar (2009) provides a survey of the literature. There is also a parallel but less extensive literature on using the yield curve as a predictor of future inflation growth; e.g Mishkin (1990).

[2] As noted in Estrella (2003), theoretical justifications for yield curve and output growth relationships range from real business cycles, countercyclical monetary policy, and life-cycle consumption (the standard first order condition used in practically all types of asset pricing models states that there is a positive functional relationship between future real consumption growth and current real interest rates, or nominal interest rates if inflation expectations embedded in interest rates are assumed to play only a secondary role). Rendu de Lint and Stolin (2003) and Estrella (2005) are examples of formal structural models.

[3] The earliest hint of a possible relationship between the slope of the yield curve and US recessions was in Kessel (1965), and later Butler (1978). Other early examples are Harvey (1988), Laurent (1988), (1989), and Stock and Watson (1988), (1989) began including the slope of the yield curve in the construction of coincident and leading indicators for the US economy.

than a univariate benchmark based on GNP only. Baltzer and Kling (2007) find similar results using German data from 1870 to 2003. However, both these articles and other analysis such as Estrella et al. (2003) and Giacomini and Rossi (2006) find that the reliability of the forecasts vary materially with time. A very powerful testament to the forecasting power of the yield curve in real time is Rudebusch and Williams (2008). It finds that the yield curve slope has been a better predictor of US recessions than the Survey of Professional Forecasters from when the latter began in 1968, and even from 1990 when forecasters should have become aware of the leading indicator property of the yield curve.

Another branch of the recent literature examines the dynamics of yield curve data in conjunction with macroeconomic data. For example, Diebold et al. (2006) establish within a VAR framework how output growth and inflation influence future movements in the yield curve and vice-versa. More sophisticated models with affine arbitrage-free models of the term structure also condense to VAR specifications in yield curve and macroeconomic data, and find similar results. Examples are Dewachter et al. (2006), Hordahl et al. (2008), and Ang et al. (2006). An interesting finding of the latter article is that the level of the short rate has better forecasting power than the yield curve slope over the 1990-2001 out-of-sample period.

To our knowledge only two papers have been written on the topic for New Zealand, with both testing the standard in-sample OLS relationship between GDP growth and the lagged yield curve slope. Guender and Moersch (1994) find no evidence for the relationship, although the paper notes the influence of the changing economic environment over the 1977 to 1992 sample period (e.g. financial market deregulation). Using data from 1988 to 2007, Wu (2009) finds statistically significant relationships analogous to the international literature.

Our forecasting investigation applies various specifications of the standard single-equation OLS regressions already noted, and a variety of simple VAR models with yield curve and GDP data that are suggested by more recent applications.

We first cover the in-sample relationships, mainly for comparison to the earlier literature. We then focus on using the various models for testing the out-of-sample forecasting power of yield curve data, which has been less-investigated internationally (and not at all in the New Zealand context). In particular, we forecast using the data strictly as it was available in real time. That is, we allow for the timeliness of the yield curve data (i.e it is always available and up to date) and the lagged reporting and historical revisions to GDP data (i.e it has a publication lag of at least one quarter, and can undergo material revisions on subsequent GDP releases). Using real time data is essential when assessing forecasting power. Many papers have documented that not taking the real time aspect into account will overestimate the forecasting power of the model(s) (Stark and Croushore (2001)).

As an extension of the out-of-sample investigation we test a simple model combination strategy. Model combination is a much-used method in modern forecasting applications and can potentially lead to enhanced forecasting performance and mitigate problems of individual model instabilities. From the perspective of forecasting economic growth using yield curve information, potential instabilities are suggested by the international evidence noted earlier; i.e the variability in out-of-sample slope-based forecasts and the potential role of the outright level of interest rates in some periods.

Finally, as a logistical innovation relative to previous studies that involve New Zealand yield curve data, we use standardised data generated by fitting the Nelson-Siegel (NS) yield curve model. The fitted data allows for the testing of many different spread measures, is less affected by anomalies in the raw data that can arise for reasons not linked to fundamentals or future expectations (e.g. liquidity issues) and the NS fitted components can also be used directly as data (as in Diebold et al. (2006).

The rest of this article is organised as follows. Section 2 outlines the individual model specifications and methods. Section 3 describes the data that we use within the forecasting framework and section 4 presents the in-sample estimation results for comparison to the existing literature. Section 5 proceeds to the evaluation of the real-time out-of-sample forecasting performance, including a comparison of our results to a set of alternative models. Section 6 describes and reports the model combination framework and results, while section 7 concludes with an overview of the results and their implications, including potential future work.

2 The forecasting framework

This section outlines the various OLS and VAR models we use, their estimation, and how each model is used to estimate in-sample relationships and to generate out-of-sample forecasts. The discussion on the model combination strategy is left to section 6.

The use of OLS and VAR models is common in the forecasting literature, given their simplicity has typically proven hard to beat. One common feature for all the models used is parsimony, which we impose based on the well known principle that such models usually forecast better than models with many estimated parameters.

Hence, no OLS or VAR specification uses more than one lag of any variable or more than four explanatory variables.

Another feature of all the models is that, when they are used for out-of-sample forecasting, we ensure that data used to estimate them is strictly that which would have been available at the time. We use the terminology "real-time out-of-sample" (RTOOS) forecasts and additional notation to distinguish between the in-sample and RTOOS specification. The discussion of the data itself follows in the subsequent section.

2.1 The OLS specification

The single-equation OLS specifications are the most common in literature and are defined as follows:

$$g_t = \alpha^h + \beta^h X_{t-h} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d}(0, \sigma_\varepsilon^2), \tag{1}$$

where g_t is quarterly GDP, h is the forecasting horizon, α^h and β^h are horizon-specific OLS parameters, and ε_t is the residual. X_{t-h} is a vector of explanatory variables, potentially including up to four variables:

$$X_{t-h} = [g_{t-h}, (\gamma_{t-h+p}^{\tau l} - \gamma_{t-h+p}^{\tau s}), \gamma_{t-h+p}^{\tau s}, \theta_{t-h+p}], \tag{2}$$

where g_{t-h} is quarterly GDP lagged h quarters, $(\gamma_{t-h+p}^{\tau l} - \gamma_{t-h+p}^{\tau s})$ is the yield spread (i.e. the difference between the yield γ_{t-h+p} at the long maturity τl and the short maturity τs), $\gamma_{t-h+p}^{\tau s}$ is a short rate, and θ_{t-h+p} is some other explanatory variable like the Nelson-Siegel measure of the slope, all lagged $h + p$ quarters. Hereafter, we refer to this specification as OLS.

The p parameter is a switch between estimating the OLS equations as in-sample relationships and the RTOOS forecasting exercise. Further details are provided in the respective results sections, but in brief: $p = 0$ for the in-sample estimation; and $p = 1$ for the RTOOS forecasting exercise (because GDP for a given quarter is published with considerable lags in New Zealand, while the yield curve data is already available).

2.2 The VAR specification

We specify the reduced form VAR models as:

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d}(0, \Sigma). \tag{3}$$

where y_t is a vector potentially containing between two and four variables, depending on the model specification. That is:

$$y_t = [g_t, (\gamma_t^{\tau l} - \gamma_t^{\tau s}), \gamma_t^{\tau s}, \theta_t], \tag{4}$$

where the variables in y_t are as defined in equation 2, but without the lagged specification (given the lag structures are allowed for in the VAR specification and the Kalman filter below). In the rest of the paper we refer to this specification as the VAR specification. Each equation in the model is estimated by ordinary least squares.

Typically, forecasting with a VAR proceeds by simply iterating forward the last complete point of data used in the information with the estimated coefficient matrix. However, such forecasts would not take account of the most up-to-date yield curve information and so would not be an RTOOS forecast.

We therefore propose a simple state space model set up of the VAR specification, which allows us to condition the growth forecast on the latest yield curve information.

$$\widetilde{y}_t = \Lambda y_t, \tag{5}$$

$$y_t = \alpha + \beta y_{t-1} + R\xi_t, \tag{6}$$

where Λ and R are diagonal matrices and y_t, α and β are as specified above. Equation 5 is the observation equation, and equation 6 is the transition equation of the system. The residuals ξ_t in the transition equation are assumed to be i.i.d, with mean zero and covariance H. The parameters in α and β and the covariance matrix H are taken from the VAR model estimation.

Applying the Kalman Filter to this state space system enables us to update the state vector at any time t, even if there are missing observations in the observation equation of the system. The gain that can be made by using the state space set up comes from the size of the Kalman gain, which determines the weight assigned to new information about y_t contained in the prediction error $\widetilde{y}_t - \Lambda y_{t|t-1}$.[4] We treat the updated Kalman Filter value as the first step forecast from this model.

[4] See Harvey (1989) for details.

3 Data

This section outlines the yield curve data and GDP data used in the analysis.

3.1 Yield curve data

The yield curve data we use in our analysis is primarily the Nelson-Siegel (NS) yield curve data generated from fitting the government yield curve data. This data set contains zero-coupon continuously-compounding rates for exact maturities of 1, 2, 3, 4, 5, 6, and 9 months, and 1, 2, 3, 4, 5, and 10 years. This range of data allows for testing many different spreads,[5] all on a consistent basis for all of the data, and without the effect of anomalies in the original data that can arise for reasons not linked to fundamentals or future expectations (e.g. occasional liquidity issues that arise in the relatively small New Zealand market, particularly because of the high proportion of bonds held by offshore investors). Appendix A details the creation of the NS data set and also the issues with alternative data sets.[6].

All of the data above are available for each business day from 1 April 1992. We transform the data to monthly observations by taking the mean over the month. Quarterly observations are recorded as the last month of the quarter. This a priori decision was made because the latest yield curve data should contain more up-to-date information about future economic growth than yield curve data earlier in the quarter.[7]

The resulting yield curve data at a quarterly frequency starts in 1992:Q2, and runs to 2009:Q1. We then construct 6 spread measures, computed as the difference between a long-maturity and a shorter-maturity rate, from these data,

$$spread_t = (100)[\gamma_t^{\tau l} - \gamma_t^{\tau s}], \tag{7}$$

where γ_t is the yield at time t, τl ranges between 4 and 10 years, and τs is either 1 month or 3 months.

[5] Ex-ante, it is impossible to know which yield and/or yield spread on the curve should be the best predictor of future growth.

[6] Because the NS data is derived, it in principle raises the issue of generated regressors. However, the very close fit of the NS data to the observed data on a daily basis (apart from anomalies, which the fitting seeks to exclude) means that the NS data can essentially be treated as observable. NS and related data is commonly used as such in practice internationally, such as the database of United States data provided by Gürkaynak et al. (2008)

[7] Nevertheless, the results using different aggregation schemes (e.g. just the end-of-month data points, or the full quarter averages) were not materially different. However, using the most up to date information tended to produce slightly better forecast performance. The alternative results are not reported, but can be provided on request.

In table 1, we report the descriptive statistics for the Nelson-Siegel fitted yield curve data. A summary of the data shows that:

- Yield dynamics are persistent, and the long-maturity rates are more persistent than the short-maturity rates;
- The short end of the yield curve is more volatile than the long end; and
- The average slope of the yield curve is zero.

The first two aspects are consistent with international statistics on yield curve data, but the third is an anomaly given that slopes of average yield curve are typically increasing and concave.[8]

3.2 Economic growth data

The economic growth data used in the estimation of the models are derived from the (log) levels of seasonally-adjusted real production GDP.

For estimating the in-sample relationships, we use the last vintage of GDP data, from which the cumulative changes for each horizon are calculated, i.e:

$$g_t = (100/h)[log(GDP_t/GDP_{t-h})],$$

where h is the forecast horizon. The cumulative change is directly comparable to most of the existing literature given it is the basis on which in-sample OLS estimations are typically undertaken. It equates to the average of the consecutive quarterly changes from $t-h$ to t. The statistics for the quarterly changes are reported at the top of table 1.

Figure 1 provides a visual representation of the in-sample data using annual GDP changes (i.e $h = 4$) and the yield curve slope calculated as the five year yield minus the 3-month rate lagged four quarters. The descriptive yield curve statistics from table 1 with an average slope of zero are apparent.

For the RTOOS application, we use the official GDP data available at the time of the forecast. The real-time data we use is from the Reserve Bank of New

[8] This anomaly in the New Zealand yield curve is well known but not well understood. One initial explanation was that it was sample specific, given the Reserve Bank was establishing its low-inflation credibility over the 1990s. However, that has proven unconvincing because the average slope of the yield curve has also been flat on average over the past decade. Another potential explanation is that (lower) global interest rates have a larger influence on New Zealand's longer-maturity rates than on shorter-maturity rates (which are more influenced by the OCR). However, that would in turn require a convincing explanation for market segmentation.

Table 1. Summary data statistics

	mean	var	skewness	kurtosis	$\hat{\rho}(1)$	$\hat{\rho}(4)$	$\hat{\rho}(8)$
GDP							
GDP	0.77	0.62	-0.25	2.65	0.30	-0.06	-0.27
Yields							
30d	6.64	2.15	0.34	2.45	0.81	0.27	0.12
60d	6.62	2.04	0.34	2.43	0.81	0.26	0.14
90d	6.60	1.95	0.34	2.42	0.81	0.26	0.15
1y	6.50	1.41	0.38	2.58	0.80	0.23	0.25
2y	6.48	1.07	0.49	2.96	0.76	0.21	0.33
3y	6.49	0.91	0.57	3.15	0.73	0.22	0.39
4y	6.52	0.84	0.63	3.17	0.71	0.23	0.42
5y	6.53	0.81	0.67	3.12	0.70	0.24	0.44
10y	6.58	0.81	0.73	2.97	0.69	0.25	0.44
Spreads							
1y90d	-0.10	0.16	0.05	2.79	0.67	0.12	-0.20
2y90d	-0.12	0.45	0.09	2.35	0.72	0.14	-0.09
3y90d	-0.10	0.70	0.12	2.09	0.75	0.14	-0.02
4y90d	-0.08	0.90	0.14	1.97	0.77	0.14	0.02
5y90d	-0.06	1.05	0.15	1.92	0.78	0.14	0.04
10y90d	-0.01	1.46	0.17	1.89	0.80	0.14	0.07
Nelson Siegel estimated factors							
NSL	6.64	0.93	0.75	3.01	0.69	0.25	0.38
NSS	0.03	2.35	-0.16	1.88	0.80	0.14	0.07
NSC	-0.57	2.24	-0.18	3.07	0.64	0.04	-0.39

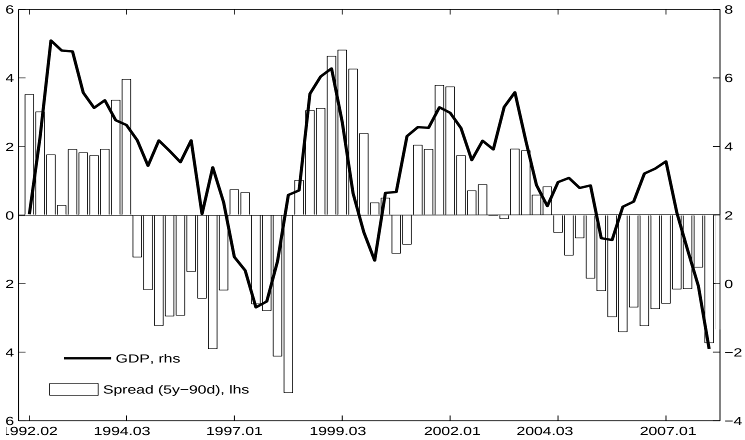
Notes: The statistics are computed on quarterly series, staring in 1992:Q2 and ending in 2009:Q1. The Nelson-Siegel (NS) estimated factors are respectively level (NSL), slope (NSS) and curvature (NSC). Daily numbers are averaged on a monthly basis. The last month of the quarter is taken as the quarterly observation. The persistency measures ($\hat{\rho}$) are sample autocorrelations computed for 1,4 and 8 quarter lags.

Zealand's real time database. We use a total of 37 vintages from this data set, spanning vintages 2000:Q1 (which contains data up to 1999:Q4) to 2009:Q1.[9] For some of the earlier vintages we have had to replace missing values. We have done this by using the growth rates from the latest available vintage. This should not affect the results, since the values missing were mostly for earlier time periods where data revisions would no longer be material.

The RMSE of the ultimate net revisions to quarterly real-time GDP data is 0.85, computed over the sample used in this analysis. That is large relative to other

[9] The vintages are for the most part in 1995–1996 NZ dollars. For the earliest vintages the prices are 1991–1992 NZ dollars.

Figure 1. GDP and the spread



Notes: GDP is measured as year on year percentage changes. The spread is the difference between the 5-year yield and the 90-day yield. The spread is lagged 4 quarters.

OECD countries.[10]

We estimate and forecast the models on quarterly log growth rates, measured as:

$$g_{t+j} = (100)[log(GDP_{t+j}/GDP_t)], \tag{8}$$

where $j = 1$. As noted in Estrella and Hardouvelis (1991), concentrating on the marginal change gives more precise information on how far into the future the term structure can predict.

4 Full-sample OLS relationships

This section reports parameter estimates of the OLS equations in section 2 over the full sample, which allows a comparison to much of the earlier literature. The results turn out to be similar to the latter, so for brevity we report just the results

[10] See Sleeman (2006) for a full description of how the real time database for New Zealand has been constructed and also for documentation on the relatively big revisions that New Zealand GDP measures undertake compared to other OECD countries. Our computed RMSE value is somewhat lower than reported in Sleeman (2006), which uses a different sample for estimation. Note that historical revisions are a material source of forecast error for the benchmark model AR(1) model in real time. When we ran the forecasting exercise with final vintage data (results available on request), the benchmark forecast errors were materially lower than for the real-time basis. These results highlight another benefit of using financial data for real-time forecasts; i.e the data are not revised.

using yield curve data composed of the 90-day short rate and the spread between the 5-year yield and the 90-day short rate. The GDP data used is the last available vintage; i.e 1992:Q2 to 2008:Q4.

In-sample relationships are estimated by setting $p = 0$ in equation 2. Thus, we do not take into account the real time information advantage of the yields. We estimate a separate OLS model for each horizon $h = 1, ..., 8$ quarters. Hence, we lose one observation for each horizon we estimate and horizons $h > 1$ induce a $h - 1$ moving average structure within the estimated residuals. We therefore use the Newey and West (1987) method to correct the standard errors for the estimated parameters, which is standard practice in the literature.

Tables 2a and 2b present the OLS estimated parameters for all of the horizons. The estimates reported in table 2a indicate that both the short rate and the spread have forecasting power for future GDP. Future economic growth is negatively related to the short rate on all forecasting horizons and positively related to the spread.

At all forecasting horizons both the short rate and the spread are highly significant. The R^2 values are always higher when using the spread. For both the short rate and the spread specification, the R^2 values drop slightly at forecasting horizons over 4 quarters and picks up again at the longest forecasting horizon.

When we expand the forecasting equations to include both the short rate and the spread, we can see from table 2a that the spread continues to be significant at the shortest and longest horizons, while the short rate loses its significance.

In table 2b we have included lagged GDP in the forecasting equations. Qualitatively, the results do not change much compared to the results in table 2a. The short rate and the spread are still highly significant at nearly all forecasting horizons, while the lagged GDP coefficient is only significant at one horizon. According to these results, the short rate and the spread actually contain more information about future GDP than GDP itself. As in the previous model specifications, it seems that including the spread produces a somewhat better in-sample fit than the short rate does.

Our estimates are comparable to the results of earlier studies mentioned in the introduction. That is, the slope of the yield curve has relatively strong explanatory power for GDP growth, but the short rate is also important (as found in Ang et al. (2006)).

Table 2. OLS parameter estimates. All forecasting horizons

(a) Short rate and spread

Forecasting horizon k quarters ahead	**OLS short rate**			**OLS spread**			**OLS short rate and spread**			
	α_0	α_1	R^2	α_0	α_2	R^2	α_0	α_1	α_2	R^2
1	1.96**	-0.71**	0.10	0.82**	1.14**	0.17	0.77	0.03	1.17**	0.17
	(0.46)	(0.27)		(0.09)	(0.32)		(0.67)	(0.41)	(0.50)	
2	1.95**	-0.69**	0.16	0.83**	1.03**	0.23	1.01**	-0.11	0.92**	0.23
	(0.38)	(0.25)		(0.09)	(0.27)		(0.48)	(0.31)	(0.41)	
3	1.84**	-0.61**	0.16	0.84**	0.84**	0.19	1.20**	-0.22	0.63	0.20
	(0.34)	(0.22)		(0.09)	(0.26)		(0.46)	(0.30)	(0.40)	
4	1.67**	-0.51**	0.14	0.84**	0.70**	0.16	1.18**	-0.21	0.49	0.17
	(0.30)	(0.19)		(0.09)	(0.24)		(0.44)	(0.29)	(0.39)	
5	1.52**	-0.41**	0.12	0.84**	0.61**	0.16	1.02**	-0.11	0.50	0.16
	(0.25)	(0.15)		(0.08)	(0.22)		(0.38)	(0.25)	(0.37)	
6	1.44**	-0.37**	0.11	0.84**	0.54**	0.15	0.99**	-0.09	0.45	0.16
	(0.23)	(0.15)		(0.07)	(0.22)		(0.34)	(0.23)	(0.33)	
7	1.36**	-0.32**	0.11	0.84**	0.52**	0.17	0.88**	-0.03	0.49	0.17
	(0.24)	(0.15)		(0.07)	(0.22)		(0.35)	(0.23)	(0.32)	
8	1.37**	-0.33**	0.13	0.83**	0.52**	0.20	0.87**	-0.03	0.50*	0.20
	(0.23)	(0.15)		(0.06)	(0.21)		(0.33)	(0.22)	(0.29)	

(b) Including lagged GDP

Forecasting horizon k quarters ahead	**OLS short rate and lagged GDP**				**OLS spread and lagged GDP**			
	α_0	α_1	α_3	R^2	α_0	α_2	α_3	R^2
1	1.65**	-0.59**	0.18	0.13	0.81**	1.23**	0.05	0.23
	(0.50)	(0.27)	(0.12)		(0.14)	(0.34)	(0.13)	
2	1.49**	-0.54**	0.25*	0.21	0.75**	0.96**	0.11	0.25
	(0.42)	(0.23)	(0.14)		(0.16)	(0.30)	(0.14)	
3	1.70**	-0.58**	0.06	0.18	0.89**	0.87**	-0.08	0.19
	(0.47)	(0.25)	(0.16)		(0.16)	(0.29)	(0.17)	
4	1.65**	-0.52**	-0.01	0.18	0.90**	0.69**	-0.12	0.15
	(0.44)	(0.22)	(0.18)		(0.18)	(0.25)	(0.19)	
5	1.47**	-0.42**	-0.01	0.15	0.89**	0.54**	-0.11	0.14
	(0.35)	(0.17)	(0.19)		(0.18)	(0.20)	(0.18)	
6	1.35**	-0.36**	0.02	0.15	0.86**	0.44**	-0.08	0.12
	(0.28)	(0.15)	(0.19)		(0.16)	(0.18)	(0.16)	
7	1.23**	-0.31*	0.06	0.14	0.81**	0.38**	-0.04	0.13
	(0.24)	(0.16)	(0.20)		(0.16)	(0.18)	(0.15)	
8	1.24**	-0.31*	0.03	0.16	0.83**	0.38**	-0.07	0.16
	(0.23)	(0.16)	(0.20)		(0.15)	(0.18)	(0.15)	

Notes: Data is quarterly numbers, and estimation period is 1992:Q3 up to 2008:Q4. OLS specifications follow equation 1. Inside the parentheses are standard errors. These are Newey and West (1987) corrected standard errors that take into account the moving average created by the overlapping of forecasting horizons. *Significantly different from zero at the 10% level in a two-tailed test. **Significantly different from zero at the 5% level in a two-tailed test. $\alpha_0 = constant, \alpha_1 = 90d, \alpha_2 = (5y - 90d), \alpha_3 =$ lagged GDP. In the model including both the short rate and the spread, $\alpha_1 = 30d$

5 Out of sample forecasting results

This section reports the real-time out-of-sample (RTOOS) forecasting results. Our first out-of-sample forecast is made for quarter 2000:Q1 and our last evaluated forecast is for 2008:Q4. The maximum forecasting horizon considered is 8 quarters. At most we evaluate 36 vintages of out of sample forecasts.[11] We estimate the different model specifications recursively using real-time data vintages. That is, we first estimate all the models using information spanning 1992:Q3 to 1999:Q4 and then forecast 1 to 8 quarters ahead. One quarter of information (i.e. one new GDP vintage and one quarter more of yield curve data) is added to the information set before the models are re-estimated and another vintage of out-of-sample forecasts are made. This procedure is repeated until we have 36 out-of-sample forecast vintages. In total we estimate and forecast using 57 models at each vintage.

Due to the fact that GDP data are published with a considerable lag and yield curve data are timely (i.e. always available and up to date), there will always be a ragged edge in the data set we use. At most we might have two quarters more yield data than GDP data. In our analysis we have assumed that we have one quarter more of interest rate data at each vintage we forecast.

The treatment of the ragged edge within the VAR specification, via conditioning on the up-to-date yield curve data using the Kalman Filter, has already been outlined in section 2.2. For estimating the OLS RTOOS forecasts at each time t, $p = 1$ in equation 2.[12]

5.1 Evaluation criteria

The forecasting performance is measured against an autoregressive model of order 1 (i.e an AR(1) model).[13] We call this our benchmark model.

[11] For each forecasting horizon, we lose one observation at the end of the evaluation sample. The eight step ahead forecast is thus evaluated over only 29 observation.

[12] Note that our specification estimates a contemporaneous relationship between the yields and GDP at forecasting horizon $h = 1$. Alternatively, the $p = 0$ OLS specifications could have been estimated, and the up-to-date yield curve data used at forecasting horizon $h >= 2$ to make the forecasts. We tested this, and it makes an immaterial difference to the RTOOS forecasting results. However, the results were materially worse when up-to-date yield curve data were not used. These results suggest that using up-to-date yield curve data information is the difference rather than the precise specification of the forecasting equation.

[13] Quarterly changes in GDP growth are not very persistent in New Zealand. The first sample autocorrelation coefficient is only 0.3. We have also estimated a random walk in mean model (RW). The difference in forecasting power between the AR(1) benchmark and the RW model is negligible.

We evaluate the point forecasts from the different models based on their root mean square forecasting error (RMSFE). Although the models are estimated on real time data, the forecasts are evaluated against the 2009:Q1 vintage of GDP.[14] The Diebold and Mariano (1995) test for equal forecasting performance is used to determine if the term structure models perform significantly differently from our benchmark model.[15]

We estimate and forecast quarterly changes in GDP, and the models are evaluated by how good they forecast annual changes in GDP.[16]

5.2 Results

Table 3 and 4 report the performance for all 8 forecasting horizons, using different short rates, spreads, and short rates and spreads in the model specifications described in section 2. The numbers in columns 2 to 9 are model implied RMSFE values, relative to the benchmark model. Thus, numbers below 1 indicate that the specific model performs better than the benchmark model. Numbers annotated with one or two stars are significantly different from the benchmark model at the 10 percent and 5 percent significance level. The values for the benchmark model are reported as actual numbers.

All the VAR model specifications include lagged GDP. Only some of the OLS specifications include lagged GDP. The model name column in the tables describes the model specification used for each model. The pre-fix describes the estimation method, while the suffix describes the explanatory data used in the model. If more than one explanatory variable is included, we have used the following naming convention: $<$ short rate $><$ spread $><$ lagged GDP $>$, e.g. 90d10y90dDG, where "DG" labels lagged GDP growth. At the end of each table, we have summarised the performance of the different model specifications listed in each table. For example, the row labeled $< Sh. >$, is the average of the models using the short

[14] GDP vintages are typically revised up to 2–3 years after their first publication (or even longer). Ideally we would have evaluated our forecasts against a GDP vintage not affected by future revisions. Given our already short sample, this was not feasible.

[15] For forecasting horizons above one quarter, the errors will have a moving-average structure. We control for this by using Newey and West (1987) corrected standard errors when computing the Diebold and Mariano (1995) test statistics. Due to the short evaluation period available for the New Zealand data, we prefer not to place too much emphasis on the tests for significant differences in forecasting performance.

[16] The yearly changes are computed as $Y_t = (100)[log(y_t/y_{t-j})]$, where $j = 4$. The transformation from quarterly changes to year on year changes is done in real time. Yearly changes was chosen for entirely practical reasons since the other statistical forecasting models at RBNZ historically have been evaluated using this transformation.

rates, while the row labeled $<$ Sh. and GDP. $>$ is the average of the models using both the short rates and lagged GDP.

Our analysis allows us to investigate two sources of discrepancy in forecasting power. One coming from using information from different parts of the yield curve. The other coming from differences in model specification.

Firstly, the results in table 3, describing the results for the OLS class of models, seem to favour the spread as the best forecaster of future GDP. This is especially evident at forecasting horizons 2 to 4. At the longer forecasting horizons though, the results are more equal and do not discriminate clearly between either using the short rates or the spreads. In addition, there is considerable variability within each group. However, searching for the best model among the OLS specifications in table 3 always returns a model including some measure of the spread.

The results in table 3 indicate the limited value of including lagged GDP in the forecasting equations. On average the models including lagged GDP perform best at 2 out of 8 horizons, while the best model specification in absolute terms among the OLS specifications include lagged GDP at 3 out of 8 horizons. When including lagged GDP in the forecasting equation, the best results were produced by including the longest spread in the sample as well, e.g. the 10-year minus the 90-day spread. Again, the differences in forecasting performance are not big, and overall there does not seem to be much value added in including lagged GDP in the forecasting equations.

Table 4 summaries the results using the different VAR specifications. Generally these results resemble the results reported for the OLS class of models; on average, using the spread outperforms using the short rate on nearly all forecasting horizons. Further, the best model specification, irrespective of forecasting horizon, includes the spread between the 10 year rate and the 90-day rate.

Based on the results reported in tables 3 and 4, differences in model class (i.e. OLS or VAR specification) do not seem to affect forecasting performance. There is a tendency for the VAR specifications to produce slightly better long run forecasts, while the OLS specifications produce better short term forecasts when including some measure of the spread in the forecasting models.

Almost all of the OLS and VAR model specifications outperform the benchmark model. This holds across all forecasting horizons. For the OLS specifications reported in table 3, the benchmark model only performs better than the average of the OLS models at the longest forecasting horizon. The benchmark model never performs better than the average performance of the VAR models reported in table 4.

Table 3. OLS. Out of sample forecasting performance

Model name	Forecasting horizon							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
ols 30d	0.97	0.90**	0.96	0.89	0.97	0.96	0.97	0.96*
ols 90d	0.98	0.91**	0.96	0.89*	0.95	0.96	0.97	0.96*
ols 1y	1.00	0.94	0.99	0.90*	0.93	0.95	0.97	0.97
ols 2y	1.01	0.97	1.02	0.94	0.95	0.97	0.98	0.98
ols 1y90d	0.94	0.92	0.98	0.97	1.07	1.02	1.00	0.96
ols 2y90d	0.94	0.88*	0.94	0.91	1.03	0.99	0.98	0.95
ols 3y90d	0.94	0.87**	0.91	0.87	0.99	0.96	0.97	0.94
ols 4y90d	0.94	0.86**	0.89	0.84	0.95	0.94	0.96	0.94
ols 5y90d	0.94	0.86**	0.88	0.82*	0.93	0.93	0.96	0.94
ols 10y90d	0.95	0.85**	0.87	0.79*	0.89	0.90	0.95	0.95
ols 30d1y90d	0.96	0.93	0.97	0.92	1.00	0.98	0.98	0.96
ols 30d2y90d	0.96	0.91	0.94	0.90	1.00	0.98	0.99	0.95
ols 30d3y90d	0.96	0.90	0.92	0.87	0.99	0.98	0.99	0.95
ols 30d4y90d	0.96	0.89*	0.91	0.85	0.97	0.97	1.00	0.95
ols 30d5y90d	0.96	0.88*	0.90	0.84*	0.95	0.96	1.00	0.95
ols 30d10y90d	0.97	0.88**	0.88	0.81*	0.91	0.94	1.00	0.96
ols 30dDG	0.97	0.94	0.93	0.92	0.98	0.97	0.95	1.05
ols 90dDG	0.97	0.94	0.94	0.92	0.97	0.96	0.95	1.05
ols 1yDG	0.98	0.97	0.96	0.93	0.95	0.95	0.95	1.06
ols 2yDG	0.99	0.99	1.00	0.98	0.97	0.97	0.97	1.07
ols 1y90dDG	0.96	0.95	0.94	0.99	1.08	1.04	0.99	1.09
ols 2y90dDG	0.96	0.92	0.89	0.93	1.04	1.00	0.96	1.07
ols 3y90dDG	0.96	0.91	0.86	0.88	0.99	0.97	0.95	1.06
ols 4y90dDG	0.96	0.90	0.84	0.85	0.96	0.95	0.95	1.06
ols 5y90dDG	0.96	0.89	0.83*	0.84	0.94	0.94	0.94	1.06
ols 10y90dDG	0.97	0.89	0.82*	0.80*	0.89	0.91	0.94	1.07
ols 30d1y90dDG	0.98	0.96	0.93	0.94	1.01	1.00	0.96	1.09
ols 30d2y90dDG	0.99	0.95	0.90	0.92	1.01	0.99	0.96	1.07
ols 30d3y90dDG	1.00	0.95	0.87	0.90	1.00	0.98	0.97	1.06
ols 30d4y90dDG	1.01	0.96	0.86	0.87	0.98	0.97	0.98	1.06
ols 30d5y90dDG	1.01	0.96	0.85	0.86	0.96	0.97	1.00	1.07
ols 30d10y90dDG	1.01	0.95	0.83	0.83	0.91	0.96	1.03	1.10
Averages								
All	0.97	0.92	0.91	0.89	0.97	0.97	0.97	1.01
Sh.	0.99	0.93	0.98	0.91	0.95	0.96	0.97	0.97
Sp.	0.94	0.87	0.91	0.87	0.98	0.96	0.97	0.95
Sh. and Sp.	0.96	0.90	0.92	0.87	0.97	0.97	0.99	0.95
Sh. and GDP	0.97	0.94	0.94	0.92	0.98	0.97	0.95	1.05
Sp. and GDP	0.96	0.91	0.86	0.88	0.98	0.97	0.96	1.07
Sh.,Sp and GDP	1.00	0.96	0.87	0.89	0.98	0.98	0.98	1.08
Benchmark model								
AR	0.91	1.25	1.66	1.74	1.71	1.67	1.68	1.70

Notes: All model numbers are relative RMSFE values to an AR(1) benchmark model. The numbers for the benchmark model are actual RMSFE numbers for each horizon. The evaluation period starts in 2000:Q1 and ends in 2008:Q4 for the one step forecasts. Significant differences in forecasting performance relative to the benchmark are tested by the Diebold and Mariano (1995) method. *Significantly different from zero at the 10% level in a two-tailed test. **Significantly different from zero at the 5% level in a two-tailed test. The model names are displayed using the following conventions for the individual models: $<$ short rate $><$ spread $><$ lagged GDP $>$, eg. 90d10y90dDG, where "DG" labels lagged GDP, and following these conventions for the averages: $<$ *Sh.* $>$=average of the models using the short rate, $<$ *Sp.* $>$=average of the models using the spread, $<$ Sh. and Sp. $>$=summary of models using both the short rate and the spread, and finally $<$ Sh., Sp. and GDP $>$=average of the models using the short rate, the spread and lagged GDP.

Table 4. VAR. Out of sample forecasting performance

Model name	Forecasting horizon							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
var 30d	0.96	0.93*	0.92	0.90	0.94	0.94	0.95	0.96
var 90d	0.96	0.93*	0.93	0.90	0.93	0.93	0.95	0.96
var 1y	0.96*	0.95*	0.95	0.92	0.92	0.94	0.96	0.98
var 2y	0.97**	0.97	0.98	0.96	0.96	0.97	0.99	1.01
var 1y90d	0.98	0.95	0.95	1.00	1.06	1.00	0.99	0.97**
var 2y90d	0.97	0.92	0.90	0.93	1.00	0.96	0.96	0.94
var 3y90d	0.97	0.90	0.87	0.89	0.96	0.93	0.93	0.92
var 4y90d	0.96	0.89	0.86*	0.86	0.92	0.91	0.92	0.91
var 5y90d	0.96	0.89	0.85*	0.84	0.90	0.90	0.91	0.90
var 10y90d	0.97	0.89*	0.84**	0.82	0.87	0.88	0.89	0.89
var 30d1y90d	1.03	0.99	0.96	0.96	0.96	0.93	0.95	1.00
var 30d2y90d	1.04	0.98	0.93	0.93	0.95	0.93	0.96	1.02
var 30d3y90d	1.04	0.96	0.89	0.90	0.94	0.93	0.96	1.02
var 30d4y90d	1.04	0.95	0.87	0.88	0.92	0.93	0.96	1.02
var 30d5y90d	1.03	0.94	0.86	0.87	0.91	0.92	0.96	1.01
var 30d10y90d	1.01	0.93	0.85*	0.85	0.89	0.91	0.94	0.99
Averages								
All	0.99	0.94	0.90	0.90	0.94	0.93	0.95	0.97
Sh.	0.96	0.95	0.95	0.92	0.94	0.95	0.96	0.98
Sp.	0.97	0.91	0.88	0.89	0.95	0.93	0.93	0.92
Sh. and Sp.	1.03	0.96	0.89	0.90	0.93	0.93	0.96	1.01
Benchmark model								
AR	0.91	1.25	1.66	1.74	1.71	1.67	1.68	1.70

Notes: See table 3 for explanations.

Many term structure models also provide significantly better forecasts than the benchmark model. Especially strong are the forecasting results for horizons 2 to 4, where almost 20 different models display significant forecasting performance.

Overall, the best model in our model suite is the VAR model using the longest spread available in the sample, the 10-year yield minus the 90-day rate. This model is about 18 percent better than the benchmark model at the fourth forecasting horizon, and it is significantly better than the benchmark model at the second and third forecasting horizon.

Table 5 reports the forecasting performance using the estimated Nelson-Siegel factors themselves as explanatory variables.[17] The numbers are in line with the findings reported above. The Nelson-Siegel factor resembling the slope (NSS), performs better than the other factors at all forecasting horizons. In fact, the

[17] We only report the results for the OLS model not including lagged GDP, since earlier results have shown that including lagged GDP does not help much for forecasting performance.

Table 5. VAR with Nelson-Siegel factors. Out of sample forecasting performance

Model name	**Forecasting horizon**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
ols NSL	1.01	0.99	1.02	0.97	0.99	1.01	1.04	1.01
ols NSS	0.95	0.85**	0.87	0.79*	0.89	0.91	0.95	0.94
ols NSLNSS	0.96	0.88**	0.89	0.81*	0.91	0.94	1.00	0.96
var NSL	1.02	1.01	0.99	1.00	0.99	0.99	0.99	0.98
var NSS	0.97	0.89*	0.84*	0.82	0.87	0.88	0.89	0.89
var NSLNSS	1.01	0.93	0.85*	0.85	0.90	0.92	0.94	0.98
Benchmark model								
AR	0.91	1.25	1.66	1.74	1.71	1.67	1.68	1.70

Notes: The model names are displayed using the following conventions: < estimation method >< Nelson-Siegel level >< Nelson-Siegel slope >, eg. var NSLNSS, where $< NSL >$=Nelson-Siegel level factor, $< NSS >$=Nelson-Siegel slope factor. See table 3 for further explanations.

forecasting performance of the NSS factor is comparable to the forecasting performance reported in table 3 and 4 using the spread between the 10-year yield and the 90-day yield in either the OLS or VAR model specification. This is not surprising, given the NSS slope coefficient is essentially a measure of the slope estimated over the entire yield curve. The similarity in the two measures was also apparent in the sample statistics reported in table 1.

Both of our model specifications, OLS and VAR, take advantage of the timeliness of the yield curve data. To show the advantage of doing so, we have also run the forecasting experiment without taking advantage of this extra information. That is, we ran the regressions on a balanced panel, and forecasted without conditioning on the observed yields for the subsequent quarter. These results, reported in table 6 reveal that there are large gains to be made by exploiting the extra quarter of information in the estimation and forecasting process. Performance increases by between 10 and 15 percent for forecasting horizons 2 to 4 for both the OLS and VAR class of models when explicitly taking advantage of the extra quarter of information in the different spreads used in the analysis. For the models including the short rate, the gains are not that large, but are still almost 10 percent on some forecasting horizons. On average, exploiting the extra quarter of information is better than not doing so across all forecasting horizons and across all model specifications. Exceptions are for the one step ahead forecasts when the spread is not included and forecasting horizons 7 and 8 for the OLS specifications.

Table 6. The advantage of timeliness

Model name	Forecasting horizon							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Relative difference in OLS performance								
All	1.00	0.94	0.90	0.87	0.94	0.97	1.01	1.04
Sh.	1.04	0.98	0.96	0.93	0.94	0.97	0.98	1.01
Sp.	0.98	0.90	0.87	0.84	0.94	0.96	1.00	1.04
Sh. and Sp.	0.99	0.92	0.88	0.84	0.93	0.96	1.02	1.07
Sh. and GDP	1.02	0.97	0.96	0.92	0.95	0.97	0.98	1.00
Sp. and GDP	0.99	0.92	0.87	0.85	0.94	0.97	1.02	1.05
Sh., Sp. and GDP	1.01	0.96	0.89	0.87	0.94	0.98	1.05	1.08
Relative difference in VAR performance								
All	1.01	0.97	0.93	0.92	0.96	0.97	0.97	0.99
Sh.	1.01	0.99	0.99	0.96	0.96	0.97	0.98	0.99
Sp.	0.99	0.94	0.90	0.89	0.96	0.97	0.99	0.99
Sh. and Sp.	1.04	0.99	0.92	0.92	0.96	0.96	0.95	0.98

Notes: All numbers are relative RMSFE values to the same model specifications not exploiting the extra quarter of information. Thus, numbers below 1 indicates that using extra information provides better forecasts than not using it. See table 3 for further explanations.

5.3 Comparison with other predictive models

In this section we compare the forecasting performance of the term structure models with other types of statistical forecasting models. In particular, we compare the models against the suite of models used for short term forecasting by the Reserve Bank of New Zealand. This suite consists of two BVAR models, a quarterly factor model and a VAR model. Note that all of these models use large data sets to produce their forecasts. The model labeled bigBVAR, for example, uses information from nearly 100 quarterly time series, while the factor model is based on almost 400 variables. For a full description of these models see Bloor (2009), and the references therein. We also include a comparison of the forecasting performance against the published growth forecasts produced by the Reserve Bank of New Zealand (RBNZ).

To save space we only report the comparison of the VAR model using the spread between the 10-year rate and the 90-day rate, which is the best model overall among the term structure models. The evaluation period is the same as the one that we used in the previous section. The forecasts from the alternative models and the RBNZ forecasts are real time forecasts. Table 7 reports the results.

At the one step forecasting horizon, the RBNZ forecasts and the BVAR forecasts perform 20 percent better than the benchmark model. The forecasting performance of the RBNZ and the BVARs are also significantly better than the benchmark model at the 5 percent level. At the second to fourth forecasting horizon,

Table 7. Forecasting comparison. Out of sample forecasting performance

Model name	Forecasting horizon							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Bigbvar	0.80**	0.79	0.81	0.76	0.96	1.00	1.02	1.08
Factor	0.88	**0.78**	**0.65****	**0.64***	**0.86**	0.95	0.97	1.09
Var	0.97	0.92	0.89	0.91	0.93	1.02	1.11	1.06
Bvar	0.80**	0.81	0.79	0.71	0.88	0.94	0.96	0.92
RBNZfinal	0.80**	0.79	0.83	0.81	0.99	1.01	1.04	1.01
var 10y90d	0.97	0.89*	0.84**	0.82	0.87	**0.88**	**0.89**	**0.89**
Benchmark model								
AR	0.91	1.25	1.66	1.74	1.71	1.67	1.68	1.70

Notes: See table 3 for further explanations.

the factor model significantly outperforms the other statistical models, as well as the RBNZ forecasts.

Our ex-post best term structure model performs worst of all the models at the one step horizon. Forecasting one year ahead and onwards, the term structure model becomes the best model. The finding that the spread has relatively better forecasting performance at longer horizons compared to other variables is consistent with findings in earlier studies. Estrella and Mishkin (1997), for example, argue that the spread dominates all other variables in predicting a recession two or more quarters into the future, and that the dominance increases with forecasting horizon.

Overall the term structure model is the best model on three out of eight horizons considered, while the factor model is best at four out of eight horizons. Further, the term structure model seems to have the most stable forecasting performance relative to the benchmark model across all horizons.

Given the simple structure and small information set used by the term structure model relative to the other models, we find the forecasting power comparisons reported in table 7 rather impressive. However, the forecasting performance is evaluated over a short period of time, and small changes to the evaluation sample will most likely change the ordering of the best models.

6 Model combination

In the analysis so far we have specified 38 OLS models and 19 VAR models, based on different combinations of spreads, short rates, and lagged GDP. A natural question is therefore: what specification should we choose? A reasonable approach would be to select one or several of the specifications based on the spread, which have proven to forecast well. That said, we only know this ex-post, and cannot necessarily be sure that the performance will continue in the future. For example, the literature in the introduction documented how the relationships between economic growth and different parts of the yield curve have changed over time. From a wider perspective, there is also a large literature on whether direct OLS or iterative VAR forecasts are best.[18]

Therefore, in this section we investigate a simple application of model combination; i.e combining the output of all models rather than selecting a small subset to use. This approach helps avoid debates about what is the best model, and should also mitigate any issues regarding the time-varying forecasting power of individual models. Model and forecast combination has a long history, dating back at least to Bates and Granger (1969). It is increasingly being applied in modern forecasting applications, giving reported results that are often advantageous relative to a single model.[19]

In section 6.1 we describe the combination methods we have applied, while section 6.2 reports the out of sample forecasting results using the model combination strategy.

6.1 Weights

Empirical findings indicate that for point forecast evaluation, simple weighting schemes often outperform more elaborate combination methods (which is often attributed to the difficulty of estimating model weights). We only consider equal weights and weights based on the inverse sum of squared forecast errors. These weights are simple, yet are still much used in the literature.[20]

Equal weights are simply $\frac{1}{m}$, where m is the number of models. The inverse sum of squared forecast error weights (invSSE) are derived as follows:

$$w_{j,t,h} = \frac{(1/SSE_{j,t,h})}{\sum_{i=1}^{m}(1/SSE_{i,t,h})}, \tag{9}$$

[18] See, for example, Marcellino et al. (2006).

[19] See, for example, Bjørnland et al. (2009) and Timmermann (2006) for an excellent survey of why model combination strategies may produce better forecasts on average than methods based on ex-ante best individual forecasting models.

[20] A more thorough investigation of different combination strategies would have been interesting, but we leave this for further research.

where j refers to an individual model, m is the number of models, h is the forecasting horizon, and t refers to time. SSE is the sum of squared forecasting errors,

$$SSE_{j,t,h} = \sum_{n=1}^{t}(y_t - \hat{y}_{j,t-h,h})^2, \tag{10}$$

where y_t is an observed value at time t and $\hat{y}_{j,t,h}$ is the h step ahead forecast for y_t by model j at time $t - h$. The combined forecast is,

$$\hat{Y}_{t+h} = \sum_{j=1}^{m}(w_{j,t,h}(\hat{y}_{j,t,h})). \tag{11}$$

Like the forecasts themselves, we apply the weights as we would have done in real time. Thus, at each time t we estimate a set of weights for a specific forecasting horizon. We then apply these weights when forecasting h periods ahead. The weights are horizon specific. Note that following this strategy, we lose h number of weights at the beginning of the sample. Note also that we do not allow for any training period. The weights are estimated on an expanding window, but the earlier weights will of course be very uncertain.

6.2 Combination results

Table 8 summaries the results from combining all of the term structure models with the naive method of equal weights, and the slightly more elaborate method of inverse sum of squared errors weights. The combined forecasts are evaluated over the same time period as the individual models.[21]

As table 8 makes clear, the difference between using equal weights, or weights derived by the inverse sum of squared forecasting errors is negligible.[22] The performances of the combination strategies outperform the benchmark model at all forecasting horizons except horizon 8, where the benchmark model and the combination schemes perform equally well. On both the second and third forecasting horizon, using equal weights to combine the models performs significantly better than the benchmark model. Using invSSE weights is significantly better than the benchmark model at only the second forecasting horizon.

[21] Due to the computation of the weights, we do lose observations at the beginning of the sample when estimating the invSSE weights. These have been replaced by using equal weights.

[22] Timmermann (2006) notes that when the different individual models have equal forecast variance, equal weights may be the appropriate weighting scheme.

Table 8. Combining models. Out of sample forecasting performance

Model name	**Forecasting horizon**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
combo equal	0.97	0.92**	0.90*	0.88	0.96	0.95	0.97	1.00
combo invSSE	0.97	0.92**	0.91	0.88	0.96	0.96	0.97	1.00
Benchmark model								
AR	0.91	1.25	1.66	1.74	1.71	1.67	1.68	1.70

Notes: All term structure models are combined using either equal (combo equal) or inverse sum of squared error (combo invSSE) weights. See table 3 for further explanations.

Compared to the results reported for the individual models in table 3 and 4, the combination schemes evaluated in this analysis perform worse than the best individual model, but better than the worst performing model at each forecasting horizon.

However, inspection of the time varying model weights indicates model instabilities. Figures 2a and 2b display the recursively estimated invSSE weights used for forecasting horizon 4 and 8.[23] The x-axes display the different individual models, ranging from 1 to 57 with only some highlighted by their names. The y-axes display the vintages which have been used to evaluate the out of sample forecasting performance, while the z-axes are the weights themselves.[24] Models including both the spread and the short rate, or only the short rate have seen their weights decline over the evaluation period. The weight for the OLS model including the NSL variable in figure 2a, for example, has an almost 30 percent lower weight at the end of the sample than it had at the beginning of the evaluation period.

Models including some measure of the spread have stable or improving weights. Including lagged GDP in the forecasting specifications (marked by the large "valleys" in the graphs), consistently gets lower weight as more and more vintages are added to the evaluation sample. This fact becomes especially clear in figure 2b.

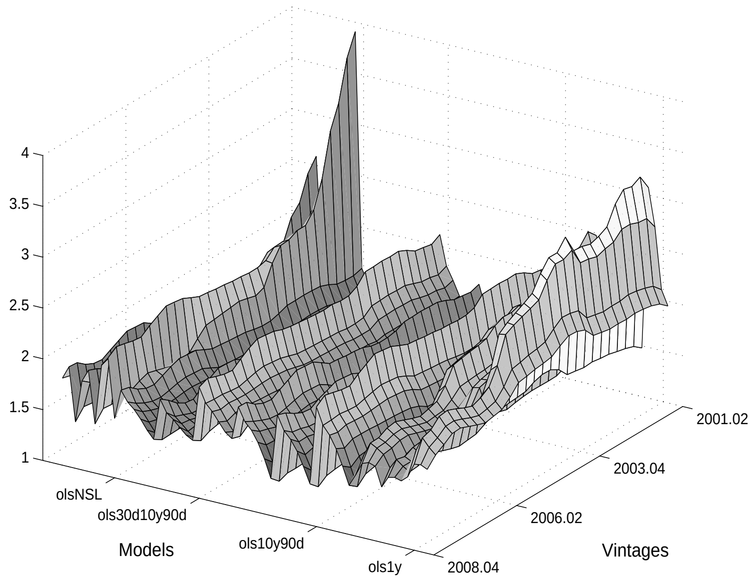
The changes in forecasting power can also be detected by looking at in-sample statistics. Figure 3 displays the difference between recursively estimated R^2 values for the OLS model including only the 90-day rate and the OLS model including only the spread between the 10-year rate and the 90-day rate. Compared to the spread model, the short rate specification had better in-sample fit at the beginning of the evaluation period, but lower R^2 towards the end of the evaluation sample.

[23] We report these two horizons because it is on these longer horizons the term structure models perform best relative to alternative models evaluated in this paper. See section 5.3.

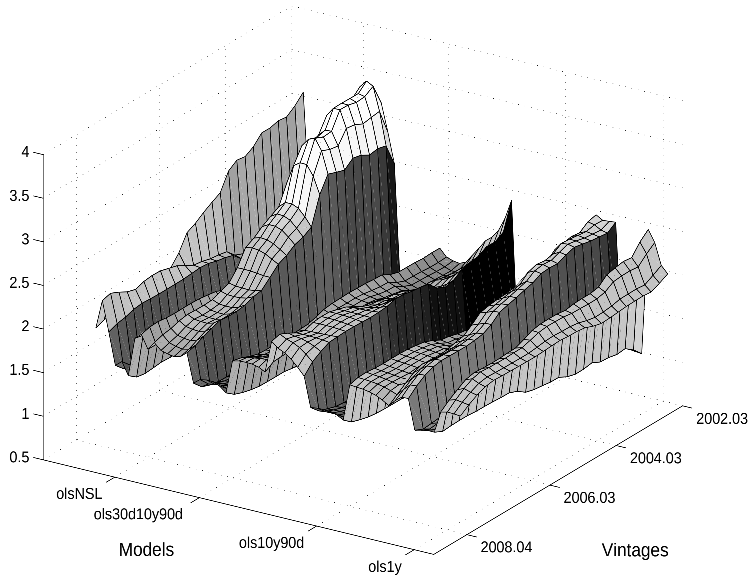
[24] The weights are multiplied by 100, and the first year of the evaluation sample, e.g. the y-axis, have been cut off because of the uncertainty of estimating the weights with few observations.

Figure 2. Recursive inverse SSE weights, 5th and 8th forecasting horizon

a) Horizon 4

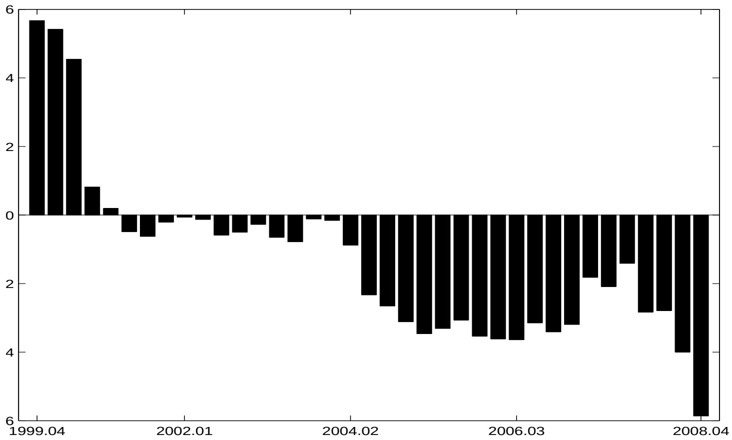


b) Horizon 8



Notes: The x axis displays the vintages used in the estimation, the y axis displays the models. Of a total of 57 individual models, only a few model names are reported in the figure. The z axis reports the estimated weights multiplied by 100. The first year of estimated weights are not displayed because of the small sample properties inherent in estimating the weights.

Figure 3. R-squared differences



Notes: The bars display the difference over time between the estimated R^2 for the OLS model including the 90-day short rate only, and the OLS model including the 10-year 90-day spread only. The estimation sample is expanding. R^2 values are multiplied by 100.

Our model combination approach reduces the risk of selecting a bad model. Evaluated over the whole sample, our model combination strategy performs better than the best model at the beginning of the evaluation sample, as the results in table 3 and 8 and figures 2 show. Continuously tracking model performance and computing relative forecasting performance (e.g. deriving weights) is a formal way of assessing stability issues in real time. Thus, the model combination strategy mitigates the issues of time varying forecasting power of individual models. Further, the model combination strategy confirms findings in the earlier literature that the relationship between the spread and GDP growth is not necessarily stable over time.

Nevertheless, our model combination results are not consistent with those of many other studies, where combinations outperform forecasts from individual models (e.g. see Koop and Potter (2004) and Clark and McCracken (2008)). There could be many reasons why our combination results do not resemble those of earlier studies. Our evaluation period is relatively short, which would probably contaminate the estimation of the weights (at least at the beginning of the sample). Further, Timmermann (2006) and Bjørnland et al. (2009) highlight how important it can be to truncate the model space.[25] We have not applied any trimming on the original model space. Thirdly, all of our models are estimated on an expanding window, so any structural breaks or changes to the data generating process would

[25] Note that the truncation of the model space, to include for example only the 10 best models before combining, can be done in real time. The difficult question is how many models to include. Only ex-post can the optimal truncation number be computed.

need time to affect the model estimates. When evaluating density forecasts, Jore et al. (2008) argue that the recursive weight strategy frequently attaches high weight to models that allow for structural change through rolling windows. Not taking this into account may hamper the forecasting performance of the model combination strategy applied in this experiment.

7 Conclusion

In this article, we have investigated the forecasting power of yield curve information for future economic growth in New Zealand.

Our in-sample OLS results parallel those of the vast literature on the topic; i.e. the slope of the yield curve is shown to have significant explanatory power for future GDP growth. We also find statistically-significant relationships between short-maturity interest rates and future GDP growth, which have been suggested in more recent literature.

Our real-time out-of-sample forecasting results reinforce the in-sample results; i.e. both short-maturity interest rates and the slope of the yield curve are better at forecasting future GDP growth than simply using GDP growth itself in an autoregressive model. In particular, exploiting the timeliness of yield curve data (it is always available and up to date) makes very material improvements to the forecasts.

Regarding model combination, the forecasting improvements often reported in this literature are not realised over our short testing period. However, our finding of variation in the predictive power of different models over the sample period nevertheless suggests that the forecast combination may still be a worthwhile exercise when applying the framework in practice.

In future research it would be interesting to bridge the purely empirical forecasting exercises from this paper to some of the more structural approaches noted in the introduction.[26] That may potentially offer an explanation for why the yield curve contains relevant information for future economic growth, and why forecasting performance changes over the evaluation sample.

From the perspective of the forecast exercises themselves, it would also be interesting to investigate in more depth the potential advantages that density combina-

[26] For example Diebold et al. (2006), which investigates the dynamic interaction between yields and macro variables with a VAR, and Ang et al. (2006), which argues a structural VAR permitting no arbitrage produces better out-of-sample forecasts than its unrestricted counterpart.

tions might offer for term structure model forecasts. However, given the relatively short testing period available for New Zealand, that investigation may be better undertaken for major economies such as the United States.

References

Ang, A, M Piazzesi, and M Wei (2006), "What does the yield curve tell us about GDP growth," *Journal of Econometrics*, 131, 359–403.

Baltzer, M and G Kling (2007), "Predictability of future economic growth and the credibility of monetary regimes in Germany, 1870-2003," *Applied Economics Letters*, 14(6), 401–404, URL http://ideas.repec.org/a/taf/apeclt/v14y2007i6p401-404.html.

Bates, J and C Granger (1969), "The combination of forecasts," *Operations Research Quarterly*, 20(4), 451–468.

Bernard, H and S Gerlach (1998), "Does the term structure predict recessions? The international evidence," *International Journal of Finance and Economics*, 3, 195–215.

Bjørnland, H C, K Gerdrup, A S Jore, C Smith, and L A Thorsrud (2009), "Does forecast combination improve Norges Bank inflation forecasts?" *Norges Bank*, *Working Paper*, 2009/01, URL http://ideas.repec.org/p/bno/worpap/2009_01.html.

Bloor, C (2009), "The use of statistical forecasting models at the Reserve Bank of New Zealand," *Reserve Bank of New Zealand: Bulletin*, 72(2).

Bordo, M D and J G Haubrich (2008), "Forecasting with the yield curve; level, slope, and output 1875-1997," *Economics Letters*, 99(1), 48–50, URL http://ideas.repec.org/a/eee/ecolet/v99y2008i1p48-50.html.

Butler, L (1978), "Recession? - a market view." Federal Reserve Bank of San Francisco Weekly Letter.

Clark, T E and M W McCracken (2008), "Averaging forecasts from VARs with uncertain instabilities," *Journal of Applied Econometrics*, (Forthcoming).

Dewachter, H and M Lyrio (2006), "A structural macro model of the yield curve," *Society for Computational Economics*, *Computing in Economics and Finance*, 236, URL http://ideas.repec.org/p/sce/scecfa/236.html.

Diebold, F and R Mariano (1995), "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–263.

Diebold, F, G Rudebusch, and S Aruoba (2006), "The macroeconomy and the yield curve: a dynamic latent factor approach," *Journal of Econometrics*, 131, 309–338.

Estrella, A (2005), "Why does the yield curve predict output and inflation?" *Economic Journal*, 115(505), 722–744, URL http://ideas.repec.org/a/ecj/econjl/v115y2005i505p722-744.html.

Estrella, A (2009), "Bibliography," http://www.newyorkfed.org/research/capital_markets/ycfaq.html.

Estrella, A and G Hardouvelis (1991), "The term structure as a predictor of real economic activity," *Journal of Finance*, 46(2), 555–576.

Estrella, A and F Mishkin (1997), "The predictive power of the term structure of interest rates in Europe and the United States: implications for the European Central Bank," *European Economic Review*, 41(7), 1375–1402.

Estrella, A, A Rodrigues, and S Schich (2003), "How stable is the predictive power of the yield curve? Evidence from Germany and the United States," *Review of Economics and Statistics*, 85(3), 629–644.

Giacomini, R and B Rossi (2006), "How stable is the forecasting performance of the yield curve for output growth?" *Oxford Bulletin of Economics and Statistics*, 68(s1), 783–795, URL `http://ideas.repec.org/a/bla/obuest/v68y2006is1p783-795.html`.

Guender, A and M Moersch (1994), "Yield Spreads and Real Economic Activity: The Case of New Zealand and Australia," *Discussion Paper (University of Canterbury, New Zealand)*, (9802).

Gürkaynak, R, B Sack, and J Wright (2008), "The U.S. Treasury yield curve: 1961 to the present," *Journal of Monetary Economics*, 54, 2291–2304.

Harvey, C (1988), "The real term structure and consumption growth," *Journal of Financial Economics*, 22, 305–333.

Hördahl, P, O Tristani, and D Vestin (2008), "The yield curve and macroeconomic dynamics," *Economic Journal*, 118(533), 1937–1970, URL `http://ideas.repec.org/a/ecj/econjl/v118y2008i533p1937-1970.html`.

Kessel, R A (1965), "The cyclical behavior of the term structure of interest rates," *National Bureau of Economic Research Occasional Paper*.

Koop, G and S Potter (2004), "Forecasting in dynamic factor models using Bayesian model averaging," *Econometrics Journal*, 7(2), 550–65.

Laurent, R (1988), "An interest rate-based indicator of monetary policy," *Federal Reserve Bank of Chicago Economic Perspectives*, 12.

Laurent, R (1989), "Testing the spread," *Federal Reserve Bank of Chicago Economic Perspectives*, 13.

Marcellino, M, J Stock, and M Watson (2006), "A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series," 135, 499–526.

Mishkin, F (1990), "The information in the longer maturity term structure about inflation," *Quarterly Journal of Economics*, 105(3), 815–828.

Mitchell, J and V S P Jore, A S (2008), "Combining forecast densities from vars with uncertain instabilities," *National Institute of Economic and Social Research, NIESR Discussion Papers*, 303, URL `http://ideas.repec.org/p/nsr/niesrd/303.html`.

Newey, W and K West (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

Plosser, C and K Rouwenhorst (1994), "International term structure and real economic growth," *Journal of Monetary Economics*, 33, 133–155.

Rendu de Lint, C and D Stolin (2003), "The predictive power of the yield curve: a theoretical assessment," *Journal of Monetary Economics*, 50, 1603–1622.

Rudebusch, G D and J C Williams (2008), "Forecasting recessions: the puzzle of the enduring power of the yield curve," *Federal Reserve Bank of San Francisco*, *Working Paper Series*, 2007-16, URL http://ideas.repec.org/p/fip/fedfwp/2007-16.html.

Sleeman, C (2006), "Analysis of revisions to quarterly GDP - a real-time database," *Reserve Bank of New Zealand: Bulletin*, 69(1).

Stark, T and D Croushore (2001), "Forecasting with a real-time data set for macroeconomists," *Society for Computational Economics*, *Computing in Economics and Finance 2001*, 258, URL http://ideas.repec.org/p/sce/scecf1/258.html.

Stock, J H and M W Watson (1988), "A probability model of the coincident economic indicators," *National Bureau of Economic Research, Inc*, *NBER Working Papers*, 2772, URL http://ideas.repec.org/p/nbr/nberwo/2772.html.

Stock, J H and M W Watson (1989), "New indexes of coincident and leading economic indicators," in *NBER Macroeconomics Annual 1989, Volume 4*, NBER Chapters, 351–409, National Bureau of Economic Research, Inc, URL http://ideas.repec.org/h/nbr/nberch/10968.html.

Timmermann, A (2006), *Forecast combinations*, vol 1, Elsevier Science B.V.

Wheelock, D and M Wohar (2009), "Can the term spread predict output growth and recessions? A survey of the literature," *Federal Reserve Bank of St. Louis Review*, September/October, Part 1, 419–440.

Wu, G (2009), "The predictive power of the term spread for future economic activity in New Zealand," *Mimeo, University of Canterbury.*

Appendix

A Appendix: Nelson and Siegel data

This appendix provides details of the yield curve data sets available for the analysis in this article and explains why the NS data is deemed to be most appropriate.

Section 1 describes the representative bank bill and government bond data already available from the Reserve Bank of New Zealand's website, section 2 describes the creation of a comprehensive government yield curve data set, and section 3 describes the creation of the fitted government yield curve data set derived from the NS model.

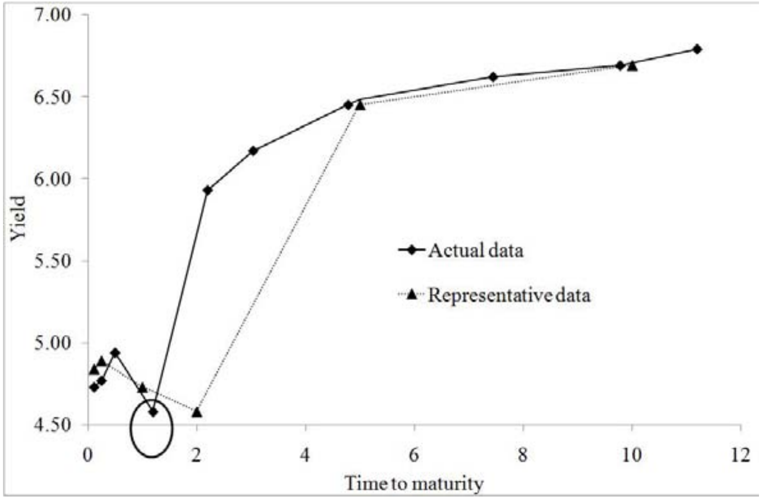
A.1 Representative data

The representative yield curve data available from the Reserve Bank of New Zealand (Bank) are the 1, 2, and 3-month bank bill rates, and the 1, 2, 5, and 10-year bond yields. The latter data have been constructed over time using the yield of a bond that has a time to maturity "close" to each of the representative maturities. However, from the perspective of empirical work in finance and economics, the representative yield curve data have several unsatisfactory aspects:

- The data are not homogeneous in their underlying characteristics. That is, bank bills have lower credit quality (and so higher yields) than government-issued securities (i.e Treasury bills and bonds), and relative liquidity can also vary considerably (particularly in periods of financial stress). Both of these aspects are evident in figures 6 and 7 from the next section.
- The times to maturity for the representative government bond data do not equal the actual times to maturity for the underlying government bonds from which the representative maturity yields were obtained. As indicated in figure 4, the representative data can therefore distort the true relationship between yields versus time to maturity.[27]
- The data sometimes reflect "scarcity premia" (or high repurchase costs) for bonds tightly held by a single market counterparty. This causes yields to be distorted away from what would be expected based on "usual" market conditions. Figure

[27] Indeed, after collating the government yield curve data in the following section, it was found that some data were not always the closest to the representative maturity. For example, some of the 1-year yields were taken from bonds with only a few months (or even days) to maturity.

Figure 4. Example of representative and government yield curve data as at 1 February 2002

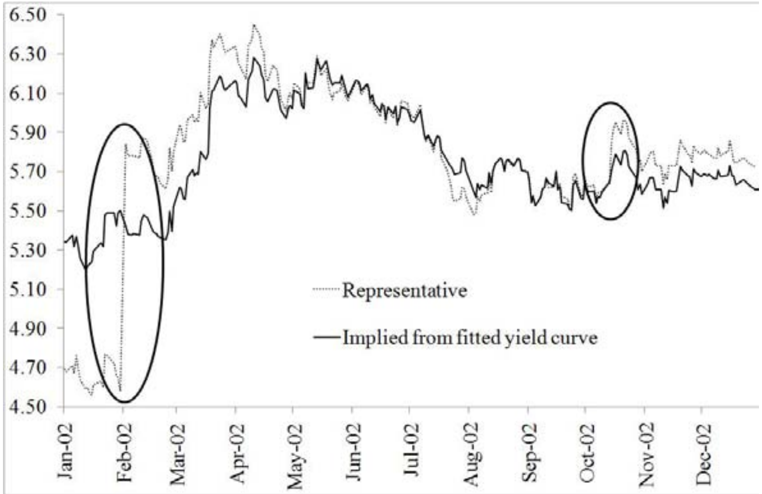


Notes: Note the distortion that arises from the representative yield curve data, and also the anomalous yield of the 15 April 2003 bond relative to the rest of the yield curve.

4 shows this for the yield curve observed on 1 February 2002, when the April 2003 bond was tightly held by an Australian participant in the New Zealand bond market.

- The data only represent a sub-set of the government bonds on issue, and so one distorted data point can have a material influence on the perceived shape of the yield curve relative to the "true" shape (as illustrated in figure 4).
- The yields are not quoted on the same basis, and are not all zero coupon (both standard criteria for use in empirical finance and economic research). The bank bill yields have a basis equal to their maturity, and the government bonds are semi-annual yields to maturity including coupon payments. Hence, the latter are effectively a weighted-average of the zero-coupon yields on the coupons and the principal, which can be materially different from the zero-coupon yield when the yield curve is not flat.
- The time series for given maturities can be discontinuous when the reference bond is switched. Figure 5 shows this effect for the representative 2-year yield, when the reference bond switched from the April 2003 bond to the April 2004 bond, and then to the February 2005 bond.

Figure 5. Example of the 2-year representative bond yield series and the 2-year fitted yield series



Notes: Changes in the representative 2-year yield occurred on 4 February (from the 15 April 2003 bond to the 15 April 2004 bond, and 16 October (to the 15 February 2005 bond).

A.2 Government yield curve data

As an initial correction to some of the unsatisfactory aspects of the representative data above, this section describes the creation of a comprehensive government yield curve data set.

The data used to define the government yield curve at each point in time are the 3-month and 6-month Treasury bill (Tbill) rates (which deals with issue of nonhomogeneous characteristics in the representative data), and the yields of all major tranches of government bonds on issue at the time along with their exact maturities (which deals with the issue of imprecision in times to maturity and the representative data being only a subset of government bond yields).

The Tbill rates are calculated as the bank bill rates for the relevant maturity less the "Tbill spread"; i.e the difference between the bank bill and Tbill rates. The Tbill spreads are obtained as:

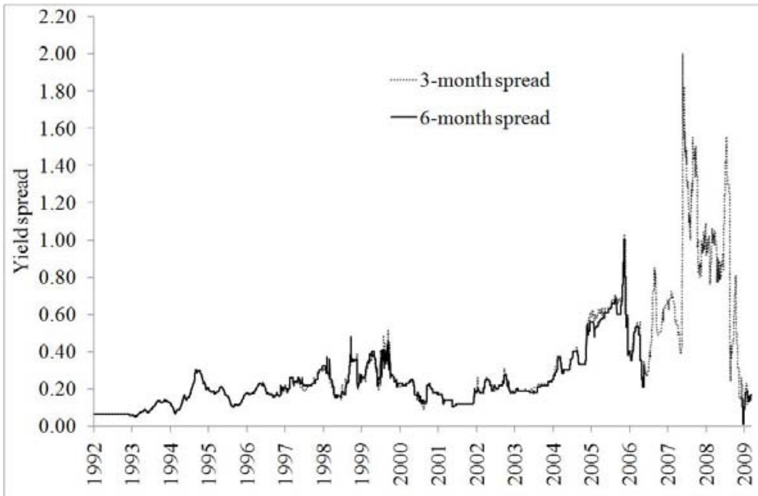
- From 20 February 1997, the spreads are those recorded by the Bank from an interbank broker. If the exact maturity is not available, the adjacent half-month spread is used (e.g 2.5-month or 5.5-month). The 6-month Tbill spread is missing between 4 Sep 2006 to 3 March 2009, because issuance was suspended during this time, and so that period therefore uses the 3-month spread as a proxy for the

6-month spread.[28]

- From 9 March 1993 to 19 February 1997, the 65-day centred moving-average from the weekly Tbill tender result less the 3-month bank bill on the day of the tender is used as a gauge for the 3-month Tbill spread. That spread is also used as a proxy for the 6-month Tbill spread.
- Prior to 9 March 1993, the spread calculated for 9 March 1993 is used. That spread is also used as a proxy for the 6-month spread.

Figure 6 illustrates the 3-month and 6-month Tbill spreads obtained as above, and figure 7 illustrates the 3-month Tbill rates calculated from the 3-month bank bill rates and the 3-month Tbill spreads in figure 3. Figure 8 illustrates the 3-month and 6-month Tbill rates.

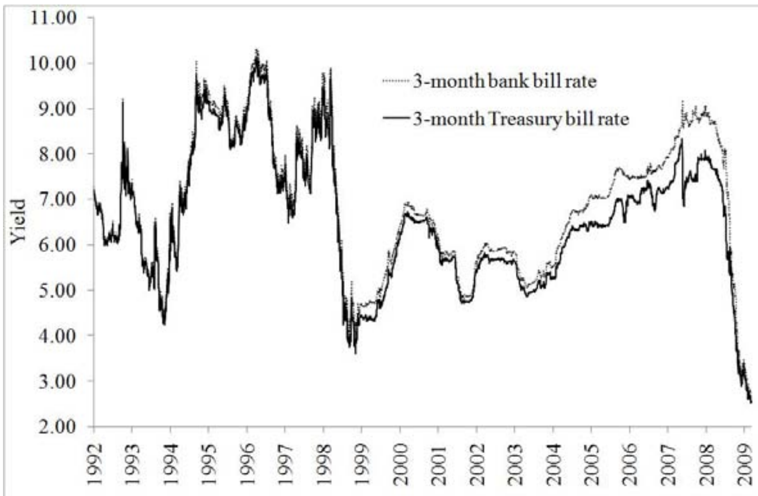
Figure 6. 3-month and 6-month Tbill spreads



[28] This is a good approximation, because figure 6 shows that the 3-month and 6-month Tbill spreads are very close to each other for the periods where independent data are available. The same comment also applies for the remaining points where 3-month Tbill spreads are used as a proxy for 6-month Tbill spreads.

Figure 8. 3-month and 6-month Tbill rates

Figure 7. 3-month bank bill and Tbill rates



The major tranches of government bonds and their characteristics are shown in the following table. The date of first issue is from records available on the New Zealand Debt Management Office website. The period of data collection for each bond refers to the data available from the Bank (the recording of secondary market data for each bond from an interbank broker typically began within a few days of the first issue date and always within a month) and the coupon rates were obtained from historical Bank records of bonds on issue.

The start date for the daily yield curve series is 1 April 1992, which is when the

Table 9. The government bonds used to define the yield curve beyond maturities of six months

First issue	Data recorded from	Data recorded to	Maturity	Coupon rate	Time to maturity at first issue	Amount issued to market
7 Apr 1988	1 Apr 1992	26 Aug 1993	15 Nov 1993	10	5.6	1469
10 Aug 1989	1 Apr 1992	15 Aug 1994	15 Feb 1995	10	5.5	1512
2 Apr 1992	1 May 1992	9 Aug 1995	15 Nov 1995	8	3.6	2475
19 Sep 1991	1 Apr 1992	25 Oct 1996	15 Nov 1996	9	5.2	2453
15 May 1990	1 Apr 1992	30 May 1997	15 Jul 1997	10	7.2	1506
21 May 1992	25 May 1992	16 Jun 1998	15 Jul 1998	8	6.1	2449
26 Aug 1993	27 Aug 1993	14 Feb 2000	15 Feb 2000	6.5	6.5	3011
10 Aug 1995	16 Aug 1995	14 Feb 2001	15 Feb 2001	8	5.5	2650
21 Mar 1991	1 Apr 1992	15 Mar 2002	15 Mar 2002	10	11.0	2512
8 Oct 1998	9 Oct 1998	14 Apr 2003	15 Apr 2003	5.5	4.5	2823
13 Aug 1992	14 Aug 1992	15 Apr 2004	15 Apr 2004	8	11.7	3044
7 Dec 2000	8 Dec 2000	15 Feb 2005	15 Feb 2005	6.5	4.2	2797
2 May 2002	3 May 2002	15 Feb 2006	15 Feb 2006	6.5	3.8	2574
11 Aug 1994	12 Aug 1994	15 Nov 2006	15 Nov 2006	8	12.3	2777
19 Aug 2004	23 Aug 2004	15 Jul 2008	15 Jul 2008	6	3.9	2700
10 Jul 1997	1 Aug 1997	n/a	15 Jul 2009	7	12.0	4247
8 Apr 1999	9 Apr 1999	n/a	15 Nov 2011	6	12.6	4416
31 May 2001	8 Jun 2001	n/a	15 Apr 2013	6.5	11.9	3947
10 Apr 2003	14 Apr 2003	n/a	15 Apr 2015	6	12.0	3234
21 Apr 2005	22 Apr 2005	n/a	15 Dec 2017	6	12.7	5317
7 May 2009	14 May 2009	n/a	15 May 2021	6	12.0	900

data for all of the major tranches on issue were simultaneously available.[29]

A.3 Generating zero-coupon data with the NS model

To correct for the remaining unsatisfactory aspects of the representative data, this section details the creation of a fitted zero-coupon dataset by applying the NS model to the government yield curve data described above.

The NS model has the following functional form for the continuously-compounding

[29] It may be possible in future to extend the dataset back further based on partial data. However, the smaller volumes issued and the lack of consistent 10-year benchmarks prior to 21 March 1991 (when the March 2002 bond was issued) would make the data less reliable in any case.

zero-coupon interest rate curve:

$$R_{\mathrm{NS}}(t,\tau) = L(t) + S(t)\left(\frac{1-\exp(-\phi\tau)}{\phi\tau}\right) + C(t)\left(\frac{1-\exp(-\phi\tau)}{\phi\tau} - \exp(-\phi\tau)\right) \tag{12}$$

where t is the time of the observation of the yield curve data, τ is the time to maturity (i.e time of maturity less time of observation), and ϕ is a parameter that governs the rate of decay of the Slope and Curvature functions.[30]

The estimation of equation 12 proceeds via the minimization of squared residuals of fitted prices over the entire sample of yield curve data.[31] That is the following function is minimized:

$$\text{Minimize} \;:\; \sum_{t=1}^{T}\sum_{k=1}^{K(t)} \left(w_{kt}\cdot\varepsilon_{kt}\right)^2 \tag{13}$$

$$\text{where} \;:\; \varepsilon_{kt} = \sum_{j=1}^{J[k]} a_{jkt}\cdot\exp\left[-\tau_{jkt}\cdot R_{\mathrm{NS}}(t,\tau_{jkt})\right] \tag{14}$$

where T is the number of yield curve observations, $K(t)$ is the number of fixed interest securities used to define the yield curve at each point in time; w_{kt} is a weighting factor;[32] $J[k]$ is the number of cashflows for security k; a_{jkt} is the magnitude of the cashflow j for security k (which is negative for the first cashflow [i.e the settlement price including accrued interest], and positive for all cashflows beyond settlement [i.e the semi-annual coupons and final coupon plus principle at maturity]); τ_{jkt} is the maturity of the cashflow j of security k; and $R_{\mathrm{NS}}(t,\tau_{jkt})$ is the NS interest rate for time to maturity τ_{jkt}.

The estimation proceeds via an iterated two-step process, i.e: (1) with starting value of ϕ, estimate the NS coefficients and the residuals for each yield curve observation in the sample; and (2) estimate a new value of ϕ via Newton-Raphson with numerical derivatives; and (3) iterate until the value of ϕ converges.

[30] The NS model has the following functional for the instantaneous forward rate curve $f_{\mathrm{NS}}(t,T) = L(t) + S(t)\exp(-\phi\tau) + C(t)\,\phi\tau\exp(-\phi\tau)$, and the usual relationship $R_{\mathrm{NS}}(t,\tau) = \frac{1}{\tau}\int_0^{\tau} f_{\mathrm{NS}}(s)\,ds$ produces the continuously-compounding zero-coupon interest rate form of the NS model.

[31] The minimization of weighted prices follows Dahlquist and Svensson (1996), and Gürkaynak et al. (2007). An alternative method for estimating the NS model is to separately calculate zero-coupon rates from the yield curve data at each point in time, and then directly regress those on the NS functions of time to maturity.

[32] This is set to the inverse of the "basis point value" (i.e the price change of the security for a yield change of a single basis point) to obtain a minimization of yield residuals. Gürkaynak et al. (2007) use a similar weighting scheme.

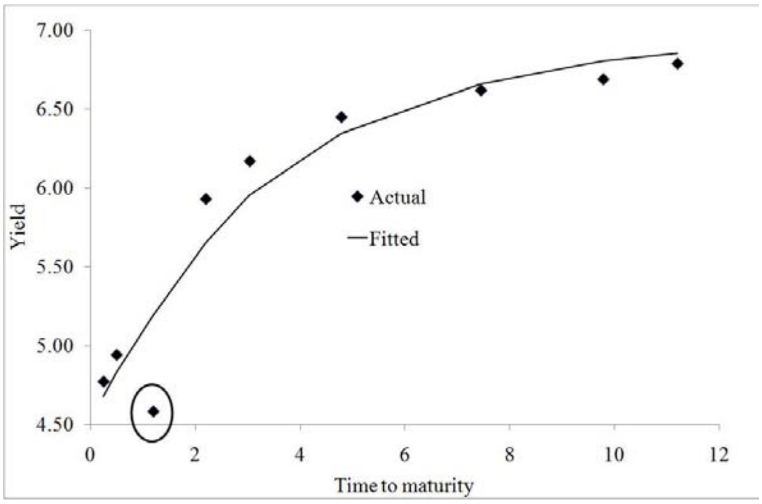
As an example of the fitted yield curve at a given point in time, figure 9 shows the results of fitting the yield curve data from 1 February 2002 shown earlier. Note how the NS fitted yield curve smoothes the anomalous yield of the April 2003 bond.

The following three figures summarize the results of the estimation. Figure 10 shows the time series of the estimated Level, Slope, and Curvature coefficients for the NS model over the full sample. Figure 11 shows examples of the yield residuals for three bonds over the sample period. Note the extreme yield residual for the April 2003 bond in 2002. As an overall gauge of the fit of the NS model to the data, figure 13 summarizes the yield residuals for each of the bonds used in the estimation (except the December 2021 that was only issued recently). The average of the 99th percentiles for the individual bonds is +/- 10 basis points.

Using the coefficients at each point in time within equation 12 creates a continuous function of time to maturity τ at that time, as already illustrated in figure 9. The zero-coupon continuously-compounding yields may then be obtained by simply evaluating that function at the desired times to maturity. The maturities selected are 1, 2, 3, 4, 5, 6, and 9 months, and 1, 2, 3, 4, 5, 7, and 10 years.

Undertaking this for the entire time series of NS coefficients then gives the time series of 1, 2, 3, 4, 5, 6, and 9 month, and 1, 2, 3, 4, 5, 7, and 10 year rates, and the relevant rates are then used in this article. For example, figure 13 illustrates the 3-month and 10-year rates.

Figure 9. Example of yield curve data as at t = 1 February 2002, and the yield curve fitted with the NS model



Notes: The fitted yields have been generated with the NS coefficients $L(t) = 7.27\%$, $S(t) = 3.74\%$, $C(t) = -1.00\%$, and $\phi = 1.013$.

Figure 10. Estimated NS coefficients

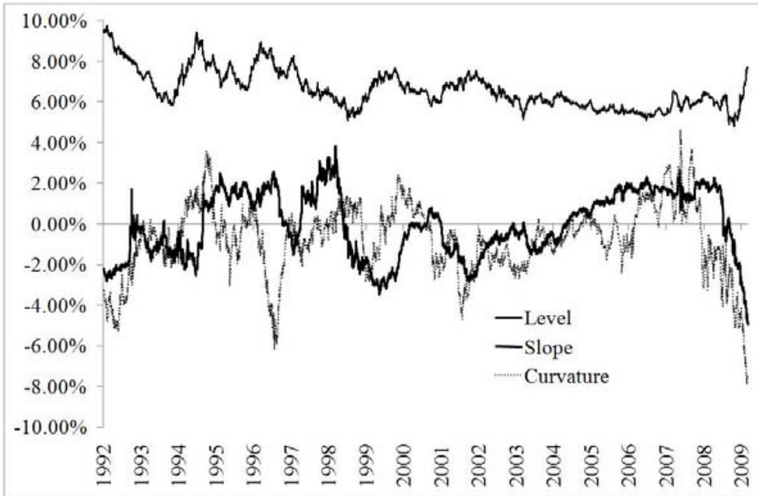


Figure 11. Yield residual examples (i.e actual yields less yields derived from the estimated NS model)

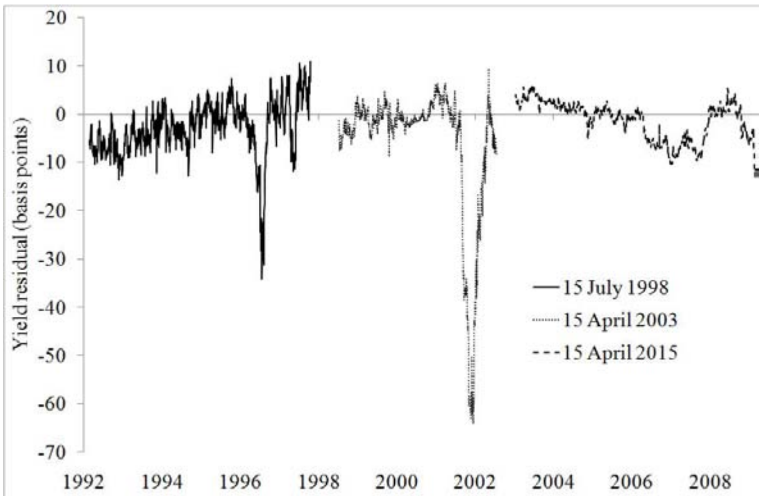


Figure 12. Yield residual summary

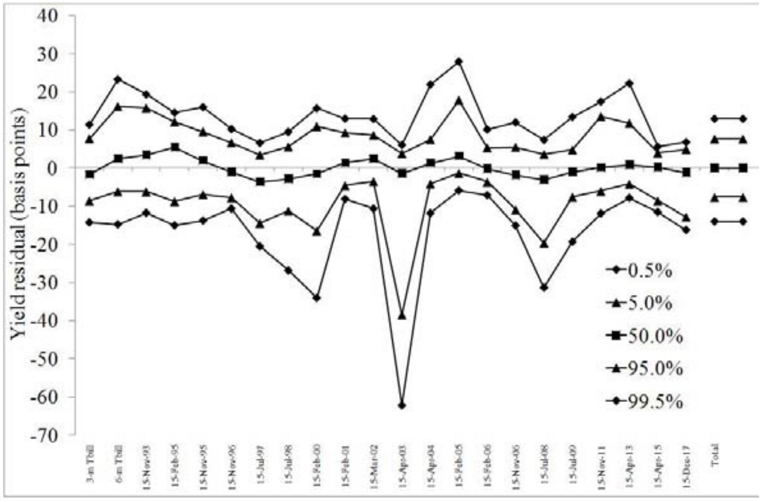


Figure 13. Example of 3-month and 10-year fitted government bond rates from the NS model, and the spread