Updating the yield curve to analyst's views

Leonardo M. Nogueira*

Banco Central do Brasil, Brazil and

ICMA Centre, University of Reading, United Kingdom

18th March, 2008

Abstract

Fixed income analysts deal constantly with the challenge of mapping their expectations about the macroeconomic environment into movements of the yield curve. This paper assumes that an analyst is able to provide a forecast of a few benchmark yields or combinations of yields. Then it derives a forecast of the entire yield curve that is consistent with the analyst's views, and computes the expected return of a bond portfolio in that scenario. We consider examples of forecasting the government bond yield curves of the United States, the Eurozone and the United Kingdom. More generally, the proposed model allows the analyst to express views on any set of correlated random variables (such as stocks, commodities, credit spreads, etc.) and to derive forecasts that are consistent with the views. The model builds on the theory of principal component analysis (PCA), can be easily extended to other markets and has no restrictions on the number of forecast variables or the number of views. A typical application is in scenario analysis, when the analyst could split the problem of forecasting the yield curve into two parts: one in which the expected developments of the macroeconomic environment are used to forecast movements of a few benchmark yields; and another part where the model derived in this paper is used to estimate the impact of the analyst's views on the entire yield curve of one country or of several countries.

Keywords: yield curves, forecasting models, scenario analysis, principal component analysis.

JEL classification: C14, C53

A long-standing problem in financial modelling is how to translate an analyst's expectations about a few market variables into reliable forecasts of other market variables. For instance, in a hypothetical scenario for the following month, a fixed income analyst has views on the short end of the US yield curve and on the long end of the UK yield curve. The analyst could then be interested in the expected movement of the Euro yield curve, the long end of the US curve and the short end of the UK curve.

*Address: Foreign Reserves Department, Banco Central do Brasil, SBS Qd. 3 Bl. B, 5º andar, Brasília, 70074-900, Brazil. E-mail: leonardo.martins@bcb.gov.br

Although cross-country yield correlations are well-known to fixed income analysts, a solution to the problem above requires forecasting a high number of variables and dealing with the complex correlation structure between different sectors of the yield curves. Besides, since the number of variables to forecast is higher than the number of views, the solution is not unique and depends on the particular method used.

One possibility is to use a multivariate regression on yield variations. In particular, because there are several variables to forecast (such as all benchmark yields of the Euro curve) a panel regression is probably necessary. However, as the number of views and the number of variables to forecast grow, the regression becomes increasingly harder to handle.

Another idea is to take a Bayesian approach and derive the conditional joint probability distribution of yields given the analyst's views — also known as the 'posterior' distribution of yields. This approach is conceptually simple, but it has the drawback that a tractable posterior distribution can be obtained only in special cases (Meucci, 2005, s. 7.1). As a result, the joint normal distribution is often used and this may be inappropriate if there is evidence of strong non-normality in the data.

A third possibility is to assume a factor model for the yield curve and try to forecast the model parameters for the next period. A good example is the popular Nelson-Siegel family of models (Nelson and Siegel, 1987; Diebold and Li, 2006). Here the factors have the nice interpretation of level, slope and curvature components of a term structure, but unfortunately calibration is non-linear and the models lose part of their tractability. Besides, these models were originally designed to work with a single term structure and extensions to multiple term structures — of several countries or different asset classes — are not straightforward (see Diebold et al. (2008) for an extension of the Diebold-Li model to multiple countries).

To overcome some of the drawbacks of the approaches above, this paper introduces an alternative, simpler model that (a) is tractable; (b) does not rely on any type of probability distribution; (c) does not assume any structure for the factors; (d) is linear; (e) is easily extended to higher dimensionality; (f) is not restricted to term structures; and finally (g) gives intuitive forecasts.

Given a set of m random variables to forecast and a set of n views on linear combinations of these variables, with $n \leq m$, we derive the expected scenario for the m random variables and assign a point estimate and a standard error to each forecast variable. This is achieved by mapping the n views into a forecast for the n most important principal components of the set of *normalised* random variables. The mapping is unique, linear and correct under the assumption that the analyst's views can be fully explained by the first n principal components.[1] That is, we assume that movements expressed in the views are caused by broad market movements (e.g. surprises about inflation, GDP growth, central bank activity, etc.) rather than by specific dynamics of individual variables — which are captured by the remaining $m - n$ principal components.

We note that the proposed model is not restricted to yield curves. It can be applied to any set of correlated random variables.[2] As it turns out, all we need to run the model is a covariance matrix and a good representation of the analyst's views. Having said that, for brevity this paper focuses on fixed income applications and considers only the case of forecasting yield curves that are consistent with views on elements of the same curves. The extension to other applications would require a straightforward change of variables.

In the following sections, first we describe our notation for the views, introduce the forecasting model and provide a simple example from the US yield curve. Next we discuss how to express uncertainty in the views. Then, we revise the example above and show, for instance, how to find the Euro yield curve that is consistent with a set of views on the US and UK yield curves. Finally, we look at the implications of these results for forecasting the return of a bond portfolio. The proof of the model is provided in the appendix.

Expressing views

Define $\boldsymbol{y}_t$ as the $m \times 1$ vector of yields[3] at time t and suppose that an analyst expresses her views on the yield curve for time $t+1$ as:[4]

$$\boldsymbol{V}\,\boldsymbol{y}_{t+1} = \boldsymbol{q}_{t+1} + \boldsymbol{\epsilon} \tag{1}$$

[1] The principal components of a random column vector $\boldsymbol{x}$ are defined as $\boldsymbol{p} = \boldsymbol{W}^T\boldsymbol{x}$, where $\boldsymbol{W}$ denotes the normalised eigenvectors of the covariance matrix of $\boldsymbol{x}$ and the elements of $\boldsymbol{p}$ are uncorrelated with each other. It follows that $\boldsymbol{x} = \boldsymbol{W}\boldsymbol{p}$ so that the first principal component explains most of the variance of $\boldsymbol{x}$, and so forth until the last principal component, which explains the least of the variance of $\boldsymbol{x}$. For an extensive review of principal component analysis, see e.g. Jolliffe (2002).

[2] For instance, suppose the analyst had three views for the next month: on realisations of the WTI crude oil price, the S&P 500 index and the iTraxx Europe index. Then, the model could be used to reveal the expected scenario for, say, the entire credit spread curve of a UK-based oil company.

[3] The vector $\boldsymbol{y}_t$ can denote any type of yield curve (e.g. zero-coupon rates, Treasury bond yields, forward rates, swap rates, etc.) or a combination of yield curves or, more generally, any set of correlated random variables.

[4] Although (1) is similar to the specification of views in the Black-Litterman portfolio optimisation model, we emphasise that, as opposed to Black and Litterman (1992), we do not take a Bayesian approach in this paper.

In (1), $\boldsymbol{V}$ is a $n \times m$ matrix that normally takes elements from the set $\{-1, 0, 1\}$, $\boldsymbol{q}_{t+1}$ is the $n \times 1$ vector of expected values of the views and $\boldsymbol{\epsilon}_{n\times 1}$ is the error in the forecast. We assume that $E[\boldsymbol{\epsilon}] = \mathbf{0}$ and $Var[\boldsymbol{\epsilon}] = \boldsymbol{\Omega}_{n\times n}$ but we make no assumption about higher moments, that is, we assume no particular distribution for $\boldsymbol{\epsilon}$. Besides, to avoid redundancy of views, we require that $rank(\boldsymbol{V}) = n$ or, equivalently, that $det\left(\boldsymbol{V}\boldsymbol{V}^T\right) \neq 0$.

The rationale for (1) is that the analyst has a forecast of where a few yields should be at $t+1$ but is not certain about the forecast, hence $\boldsymbol{\Omega}$ denotes this uncertainty. In practice, (1) could be the output of another forecasting model that links the future values of a few benchmark yields to expected movements in macroeconomic variables.

Example 1. *Suppose the analyst holds two independent views on the US Treasury bond yield curve for $t+1$:*

i. the expected 5-year yield is 5% with a standard error of 1%;

ii. the expected 2Y-10Y spread is $50bp$ with a standard error of $10bp$.

These views may be written in matrix notation as:

$$\begin{aligned}
\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} y^2_{t+1} \\ y^5_{t+1} \\ y^{10}_{t+1} \end{pmatrix} &= \begin{pmatrix} 5\% \\ 50bp \end{pmatrix} + \boldsymbol{\epsilon} \\
\boldsymbol{\Omega} = Var[\boldsymbol{\epsilon}] &= \begin{pmatrix} (1\%)^2 & 0 \\ 0 & (10bp)^2 \end{pmatrix}
\end{aligned}$$

where one can immediately identify the elements of (1).

Forecasting yields

When $n = m$ and $\boldsymbol{V}$ is invertible, (1) is a linear system whose unique solution is

$$\begin{aligned}
E\left[\boldsymbol{y}_{t+1}\right] &= \boldsymbol{V}^{-1}\boldsymbol{q}_{t+1} \\
Var\left[\boldsymbol{y}_{t+1}\right] &= Var\left[\boldsymbol{V}^{-1}\boldsymbol{\epsilon}\right] = \boldsymbol{V}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{V}^{-1}\right)^T
\end{aligned} \qquad (2)$$

However, this is hardly the case since the number of views (n) is typically less than the number of yields to forecast (m). Thus, $\boldsymbol{V}$ is not invertible and the solution to (1) is not unique in general. In fact, when $n < m$ there is an infinite number of yield curves that satisfy (1) and, to choose among all possible solutions, we propose a model that is consistent not only with the views expressed in (1) but also with the covariance matrix of yield variations.

Suppose now is time t and define $\boldsymbol{q}_t = \boldsymbol{V}\boldsymbol{y}_t$ as the *non-random* value of the views based on the yields at time t. Subtracting $\boldsymbol{q}_t$ from both sides of (1) gives

$$\boldsymbol{V}\,\Delta\boldsymbol{y} = \Delta\boldsymbol{q} + \boldsymbol{\epsilon} \tag{3}$$

where $\Delta\boldsymbol{y} \overset{def}{=} \boldsymbol{y}_{t+1} - \boldsymbol{y}_t$ and $\Delta\boldsymbol{q} \overset{def}{=} \boldsymbol{q}_{t+1} - \boldsymbol{q}_t$.

Denote the historical mean of $\Delta\boldsymbol{y}$ by $\boldsymbol{\mu}_{m\times 1}$ and the historical covariance matrix of $\Delta\boldsymbol{y}$ by $\boldsymbol{S} = \boldsymbol{DCD}$, where $\boldsymbol{D}_{m\times m}$ is the diagonal matrix of standard deviations and $\boldsymbol{C}_{m\times m}$ is the correlation matrix. These matrices are estimated at time t from historical data.

From the spectral decomposition theorem we have $\boldsymbol{C} = \boldsymbol{W\Lambda W}^T$, in which $\boldsymbol{\Lambda}_{m\times m}$ is the diagonal matrix of eigenvalues of $\boldsymbol{C}$ in descending order, and $\boldsymbol{W}_{m\times m}$ denotes the normalised eigenvectors of $\boldsymbol{C}$ in the same order as $\boldsymbol{\Lambda}$. Also define $\hat{\boldsymbol{\Lambda}}_{(m-n)\times(m-n)}$ as the sub-matrix of $\boldsymbol{\Lambda}$ with the smallest $m-n$ eigenvalues along the diagonal and decompose $\boldsymbol{W}$ into the sub-matrices $\tilde{\boldsymbol{W}}$ and $\hat{\boldsymbol{W}}$ according to

$$\boldsymbol{W} \overset{def}{=} \left[\tilde{\boldsymbol{W}}_{m\times n} \;\middle|\; \hat{\boldsymbol{W}}_{m\times(m-n)}\right]$$

Theorem 1. *Under the assumption that all yield curve movements implicit in the views can be fully explained by movements on the first n principal components of normalised yield variations, the forecast yield curve at time $t+1$ is given by:*

$$\begin{aligned} E\left[\boldsymbol{y}_{t+1}\right] &= \boldsymbol{y}_t + \boldsymbol{\mu} + \boldsymbol{DA}\left(\Delta\boldsymbol{q} - \boldsymbol{V}\boldsymbol{\mu}\right) \\ Var\left[\boldsymbol{y}_{t+1}\right] &= \boldsymbol{D}\left(\boldsymbol{A\Omega A}^T + \boldsymbol{B}\hat{\boldsymbol{\Lambda}}\boldsymbol{B}^T\right)\boldsymbol{D} \end{aligned} \tag{4}$$

with $\boldsymbol{A}_{m\times n} = \tilde{\boldsymbol{W}}\left(\boldsymbol{VD}\tilde{\boldsymbol{W}}\right)^{-1}$ and $\boldsymbol{B}_{m\times(m-n)} = \left(\boldsymbol{I}_m - \boldsymbol{AVD}\right)\hat{\boldsymbol{W}}$ where $\boldsymbol{I}_m$ is the $m\times m$ identity matrix.

Theorem 1 is the most important result of this paper. It gives a point estimate to the vector of yields $\boldsymbol{y}_{t+1}$ and the covariance matrix of this forecast. In particular, $\boldsymbol{DA} : \mathbb{R}^n \to \mathbb{R}^m$ maps the n views into a forecast of movements for m points of the yield curve, and $\boldsymbol{DB} : \mathbb{R}^{m-n} \to \mathbb{R}^m$ maps the error of the approximation using PCA (principal component analysis) into an error for the m yields. Therefore, together they are responsible for the construction of the expected scenario for the yield curve that is consistent with the views.

We note that $Var\left[\boldsymbol{y}_{t+1}\right]$ is the sum of two clearly defined terms: $\boldsymbol{DA\Omega A}^T\boldsymbol{D}$, which captures the analyst's uncertainty on the views; and $\boldsymbol{D}\boldsymbol{B}\hat{\boldsymbol{\Lambda}}\boldsymbol{B}^T\boldsymbol{D}$, which captures the error in the PCA approximation.

To use Theorem 1 we need the yield curve at time t, the set of subjective views $\left\{\boldsymbol{V}, \boldsymbol{q}_{t+1}, \boldsymbol{\Omega}\right\}$ and the mean vector and covariance matrix of yield variations. In the context of yield curves we regard $\boldsymbol{\mu} = \mathbf{0}$ as an acceptable assumption because $\boldsymbol{\mu}$ is small in general and has a secondary effect in the forecast.

Therefore, to forecast the yield curve one quarter ahead we should estimate the covariance matrix from historical quarterly yield variations, and so on. Of course this can be a problem if there is not enough data to estimate a reliable covariance matrix. In this case, it is possible to use higher frequency data and the square-root-of-time rule as long as yield variations are approximately i.i.d. (independent and identically distributed); see Alexander (2001) for a discussion of these issues.

Example from the US yield curve

The following example shows how Theorem 1 may be used to forecast the US Treasury bond yield curve that is consistent with a set of views. We consider the actual yields-to-maturity available from Bloomberg for nine benchmark maturities (1, 3 and 6-month Treasury bills; and 1, 2, 3, 5, 10 and 30-year Treasury bonds).[5]

Example 2. *Today is 31st December, 2007, when the analyst has two views on the yield curve on 31st January, 2008:*

i. the US 3-month yield is expected to decrease from 3.24% *to* 1.94%*;*

ii. the US 5-year yield is expected to decrease from 3.44% *to* 2.76%.

In matrix notation, the elements of (1) *are*

$$
\begin{aligned}
\boldsymbol{V} &= \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \\
\boldsymbol{y}_{t+1}^T &= \begin{pmatrix} y_{t+1}^{1M} & y_{t+1}^{3M} & y_{t+1}^{6M} & y_{t+1}^{1Y} & y_{t+1}^{2Y} & y_{t+1}^{3Y} & y_{t+1}^{5Y} & y_{t+1}^{10Y} & y_{t+1}^{30Y} \end{pmatrix} \\
\boldsymbol{q}_{t+1} &= \begin{pmatrix} 1.94 \\ 2.76 \end{pmatrix}
\end{aligned}
$$

[5]For instance, we use the Bloomberg code *USGG3M Index* for the 3-month yield-to-maturity. Other maturities have similar codes. Because generic yields for 1 and 3 years were not available, they are computed from the spreads relative to other maturities.

We assume that $\boldsymbol{\Omega} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{V}^T$ *where* $\boldsymbol{S}$ *is the estimated* 9×9 *covariance matrix of monthly first differences of yields from February 2003 to December 2007.*

Table 1 compares the current yield curve (on 31^{st} *December, 2007), the forecast and the realised curve on* 31^{st} *January, 2008 (i.e. what actually happened in the market). The standard errors are provided in brackets under each forecast. Exhibit 1 shows the same curves and includes the confidence intervals of yields in terms of two bands, each of them being two standard errors away from the forecast.*

	1M	**3M**	**6M**	**1Y**	**2Y**	**3Y**	**5Y**	**10Y**	**30Y**
31/12/07	2.61	3.24	3.39	3.25	3.05	3.02	3.44	4.02	4.45
Forecast (st.error)	0.87 (0.41)	1.94 (0.2)	2.27 (0.19)	2.28 (0.21)	2.10 (0.26)	2.09 (0.29)	2.76 (0.28)	3.64 (0.25)	4.31 (0.22)
Realized	1.58	1.94	2.05	2.08	2.09	2.17	2.76	3.59	4.32

Table 1: US Treasury yield curves of Example 2. The forecast of the long-term yields is accurate, but one may experience problems with short-term yields. All values are in percentages.

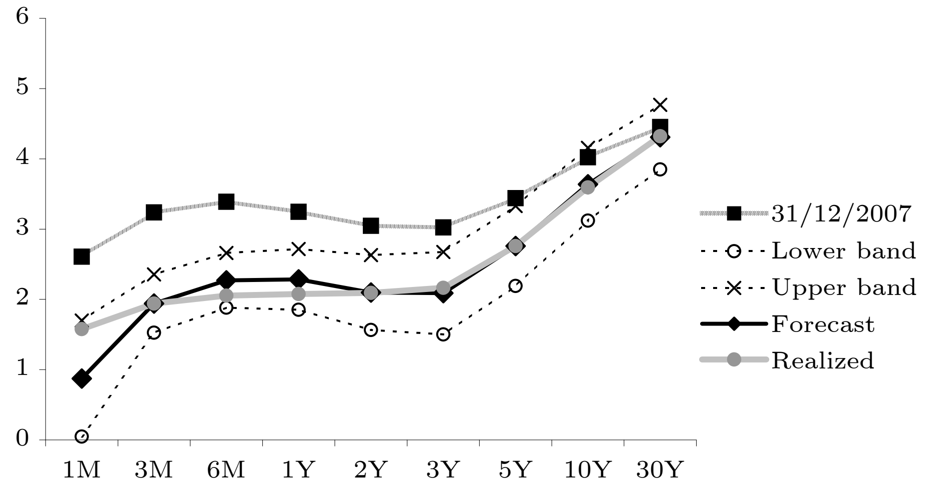

Exhibit 1: US Treasury yield curves of Example 2. The realized values are very close to the forecast for long-term yields.

The views in Example 2 were deliberately chosen to match the realised values for the 3M and 5Y yields on 31st January, 2008, and this date was chosen because yield movements were exceptionally large. As a result, yield correlations could have changed and the model could have failed. Therefore, this example allows us to answer the following question: if the analyst can provide very accurate forecasts for a few points, how good is the forecast given by Theorem 1 for the remaining points of the yield curve?

Exhibit 1 shows that the forecast is very good even during periods of extreme market activity. All realised values are within the confidence intervals given by the two bands.

Thus, providing a good forecast for the 3M and the 5Y US yields is generally enough to forecast the entire yield curve accurately.

The difference between the forecast and the realised yield curve is larger for the 1M yield. This may be due to a variety of reasons, but we emphasise that

1. The 1M yield has weak correlation with the rest of the curve. However, the effect of correlation is already taken into account by Theorem 1 so that a weaker correlation would generally imply a larger standard error in the forecast, as observed in this case.

2. The analyst has only two views, so Theorem 1 assumes that all movements expressed in the views are explained by the first two principal components alone. But since these components are responsible for the 'trend' and 'tilt' movements of the yield curve, the forecast is the combination of a parallel movement (because the views for both 3M and 5Y yields imply a negative trend) with a substantial 'steepening' of the curve (the 3M-5Y spread increased from 20bp to 82bp). As a result, both views push the 1M yield down and explain why its forecast is so low. See Loretan (1997) and Alexander (2001, ch. 6) for applications of PCA to different markets.

3. Finally, the covariance matrix may be inappropriate for a distressed period. We used the historical, equally weighted covariance matrix of monthly first differences of yields from February 2003 to December 2007. Alternatively, one could estimate this matrix using EWMA or GARCH models, for instance, because these models give higher weights to more recent information. However, the choice of estimation model is not core to this text and is left as an exercise to the reader. See Alexander (2001) for a review of models to estimate covariance matrices.

Expressing uncertainty in the views

One of the hardest tasks when expressing views in the form of (1) is to choose the uncertainty matrix $\mathbf{\Omega}$. Ideally (1) should be the output of another forecasting model that links the dynamics of a few benchmark yields to macroeconomic indicators or other observable market variables. Thus, $\mathbf{\Omega}$ would follow from the forecasting model and no further assumptions are necessary. However, when no such a model is available, one may consider one of the alternatives below.

Alternative I: One may assume that views are independent (as if drawn from independent experiments) and define $\mathbf{\Omega}$ as a diagonal matrix according to the analyst's confidence on each view:

$$\boldsymbol{\Omega}^{I} = \kappa\left(\boldsymbol{I}_n - \boldsymbol{G}\right)\boldsymbol{G}^{-1} = \kappa\begin{pmatrix} \frac{1-g_1}{g_1} & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & \frac{1-g_n}{g_n} \end{pmatrix} \tag{5}$$

where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix, $\boldsymbol{G}_{n\times n}$ is the diagonal matrix of credibility weights $g_i \in (0,1]$ and κ is an optional positive penalty term that penalises low-conviction views with an increased variance.[6] The problem with (5) is that this definition is inconsistent with empirical evidence since yields (and hence the views) are highly correlated in practice.

Alternative II: One may use $\boldsymbol{S}$, the covariance matrix of yield variations, and set $\boldsymbol{\Omega}^{II} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{V}^{T}$. This definition guarantees consistency but does not allow the analyst to express confidence on the views.

Alternative III: One may combine the alternatives below to be consistent with yield correlations and capture the analyst's confidence at the same time. This is obtained if one defines $\boldsymbol{\Omega}^{III} = \boldsymbol{H}^{T}\boldsymbol{\Omega}^{II}\boldsymbol{H}$ where $\boldsymbol{H}$ is the Cholesky decomposition of $\boldsymbol{\Omega}^{I}$, i.e. $\boldsymbol{\Omega}^{I} = \boldsymbol{H}^{T}\boldsymbol{H}$. This effectively scales up or down the variances given by $\boldsymbol{\Omega}^{II}$ according to the analyst's confidence on the views.

Alternative IV: Another idea is to allow for uncertainty on $\boldsymbol{q}_{t+1}$, which was previously defined as a known vector, and write:

$$\begin{aligned} E\left[\boldsymbol{q}_{t+1}\right] &= \boldsymbol{q}^{*}_{t+1} \\ Var\left[\boldsymbol{q}_{t+1}\right] &= \boldsymbol{\Omega}^{I} \end{aligned} \tag{6}$$

where $\boldsymbol{\Omega}^{I}$ is defined as in alternative I.[7] Next suppose that $Var\left[\boldsymbol{\epsilon}\right] = \boldsymbol{\Omega}^{II}$ in (1) and that $\boldsymbol{q}_{t+1}$ and $\boldsymbol{\epsilon}$ are independent. Then, it is easy to see that the views in (1) become:

$$\boldsymbol{V}\,\boldsymbol{y}_{t+1} = \boldsymbol{q}^{*}_{t+1} + \boldsymbol{\epsilon}^{*} \tag{7}$$

with $E\left[\boldsymbol{\epsilon}^{*}\right] = \boldsymbol{0}$ and $Var\left[\boldsymbol{\epsilon}^{*}\right] = \boldsymbol{\Omega}^{IV} = \boldsymbol{\Omega}^{I} + \boldsymbol{\Omega}^{II}$. Thus, by replacing $\boldsymbol{q}_{t+1}$ by $\boldsymbol{q}^{*}_{t+1}$ and $\boldsymbol{\Omega}$ by $\boldsymbol{\Omega}^{IV}$ in (4), the yield curve forecast follows as before.

There are certainly many other ways of expressing uncertainty in the views, but we

[6] The choice of penalty term may be linked to the risk-aversion of the analyst, for instance. Increasing the penalty does *not* change the forecast but increases the uncertainty in the forecast, i.e. the confidence interval gets wider.

[7] This approach is analogous to the Black-Litterman model, in which the analyst is uncertain about the expected asset returns to use as input for a mean-variance portfolio optimization model (Black and Litterman, 1992).

believe the alternatives above provide a reasonable starting point.

Going Global

One of the interesting consequences of Theorem 1 is that the model may be easily extended to higher dimensionality. For instance, the next example revisits our starting problem, when the analyst had views on the US and the UK yield curves and would like to forecast the impact of these views on the Euro curve.

Example 3. *Today is again 31st December, 2007, when the analyst has two views for 31st January, 2008:*

i. the US 3-month yield is expected to increase from 3.24% *to* 4.24% *with* 100% *confidence;*

ii. the UK 10-year yield is expected to decrease from 4.51% *to* 3.51% *with* 50% *confidence.*

Given these views, we wonder:

1. What is the impact of the views on our expectations for the Euro curve?

2. Should the US 10-year yield go up or down?

To answer these questions, we consider seven vertices $(3M, 1Y, 2Y, 3Y, 5Y, 10Y, 30Y)$ *for each of the three yield curves (US, Eurozone and UK) and define* $\boldsymbol{y}_{t+1}$ *as the* 21×1 *vector of yields. Next, to be consistent with Example 2, we use monthly data from February 2003 to December 2007 to estimate the mean vector and the covariance matrix of yields. We set* $\boldsymbol{\Omega} = \boldsymbol{\Omega}^{IV}$*, as in alternative IV above, and assign a penalty term of* 0.01*. Table 2 and Exhibit 2 summarise the results.*

In a scenario where the US 3M increases by 1% and the UK 10Y decreases by 1%, we have that, for instance:

- The Euro curve is expected to move down by more than 1% for long-term yields and less than 1% for short-term yields. That is, we have a 'bull flattening' on the curve.
- The UK curve also flattens but movements are smaller than in the Euro curve.
- The short end of the US curve moves up because it is highly correlated with the US 3M, but the long end moves down because the correlation is stronger with the UK 10Y.

(a) US yield curves

US	**3M**	**1Y**	**2Y**	**3Y**	**5Y**	**10Y**	**30Y**
31/12/07	3.24	3.25	3.05	3.02	3.44	4.02	4.45
Forecast (st.error)	4.24 (0.20)	3.53 (0.30)	2.82 (0.46)	2.54 (0.54)	2.50 (0.57)	2.82 (0.53)	3.21 (0.46)

(b) Euro yield curves

Euro	**3M**	**1Y**	**2Y**	**3Y**	**5Y**	**10Y**	**30Y**
31/12/07	3.76	4.07	3.96	4.04	4.12	4.33	4.60
Forecast (st.error)	3.73 (0.09)	3.75 (0.21)	3.19 (0.36)	3.05 (0.40)	2.98 (0.42)	3.18 (0.37)	3.49 (0.30)

(c) UK yield curves

UK	**3M**	**1Y**	**2Y**	**3Y**	**5Y**	**10Y**	**30Y**
31/12/07	5.45	4.46	4.36	4.34	4.41	4.51	4.30
Forecast (st.error)	5.60 (0.11)	4.29 (0.24)	3.76 (0.36)	3.57 (0.38)	3.48 (0.38)	3.51 (0.35)	3.40 (0.25)

Table 2: Yield curves on 31 Dec 2007 and forecast yield curves for 31 Jan 2008 in the US, Eurozone and UK according to the views of Example 3. Here we assume that the US 3M moves up by 1% and that the UK 10Y moves down by 1%. The Euro curve moves down because it is more correlated with the UK 10Y than with the US 3M. But, as expected, shorter-term maturities of all three curves show a good degree of correlation with the US 3M.

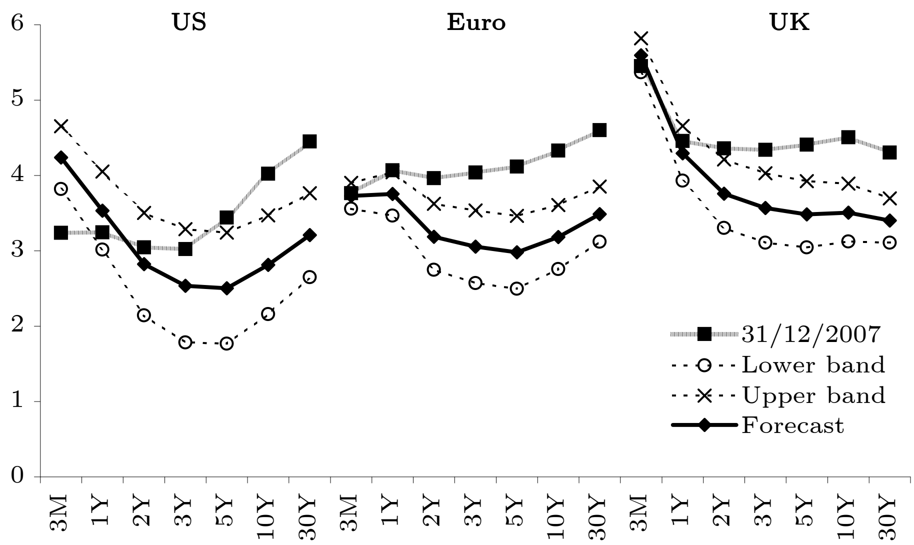

Exhibit 2: Government bond yield curves of Example 3. The movement of the UK 10Y dominates, even for long-term US yields. The confidence intervals suggest a long duration bet for the Eurozone. There is also a possibility of betting on the 'flattening' of any of the curves.

- As expected, movements of the US 3M are more relevant to short-term yields and movements of the UK 10Y are more relevant to longer-term yields. Thus, the model captures correctly the trader's intuition of how yields actually behave.

Clearly the most important information is that the UK 10Y yield is expected to move down by 1%, despite the low confidence on the view. This is because, according to Theorem 1, lack of confidence only adds uncertainty to the forecast. In fact, the expected value in (4) is not sensitive to the uncertainty at all. This finding may seem counter-intuitive at first sight, but it is consistent with the purpose of the model. Theorem 1 explores the correlation matrix of yields to provide, roughly, the 'most likely' scenario for the yield curves that is consistent with the views. Thus, the forecast of the US 3M is exactly 4.24% and the forecast of the UK 10Y is exactly 3.51% because these are indeed the views. Lower-conviction views simply imply wider confidence intervals relative to when conviction in the views is high.

To gain more intuition on the correlation between the three yield curves, Exhibit 3 plots the first three eigenvectors of the historical correlation matrix of the 21 yields. The first eigenvector, which explain 64% of the total variance of the data, is always positive and is interpreted as the 'trend' component (i.e. the three curves move up or down together most of the time). The second eigenvector explains 10.6% of the variance and is roughly a diagonal line for each of the curves. This implies that changes in the slope of the curves are correlated. The third eigenvector explains further 9% of the variance but, contrarily to empirical evidence from PCA of single term structures, it cannot be interpreted as a 'curvature' component. Instead, according to this eigenvector some parallel changes of the US curve have little impact on the Euro curve but cause a strong steepening/flattening movement of the UK curve. The economic interpretation of this behaviour is not obvious, but this factor cannot be neglected because it explains 9% of the total variance of the data.

This highlights another advantage of Theorem 1. Because it does not assume any particular structure for the common factors driving the yield curves, the model provides a forecast of the curves that is, by construction, consistent with empirical evidence and highly adaptable to a changing economic environment. Having an economic interpretation for the factors is not necessary.

Forecasting Bond Portfolio Returns

Define $b(t, y; T)$ as the price of a bond at time t with yield-to-maturity y and maturity T. For ease of notation, for every bond $i \in \{1, 2, ..., m\}$, define $b^i(y) \equiv b(t, y; T^i)$ and

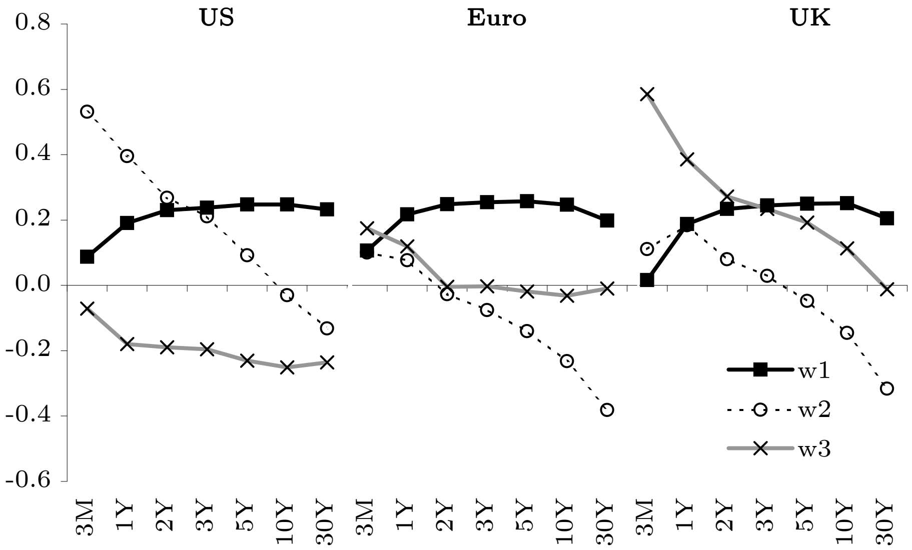

Exhibit 3: First three eigenvectors of correlation matrix of yields of Example 3. The first and second factors have the usual interpretation of 'trend' and 'tilt' components, but the third factor can no longer be interpreted as the 'curvature' component. This is because PCA has been applied to all three curves simultaneously.

$\hat{b}^i(y) \equiv b(t+1, y; T^i)$. Because $\hat{b}^i(y)$ is non-linear on y, it follows from a Taylor's series expansion of $\hat{b}^i\left(y^i_{t+1}\right)$ around $\hat{b}^i\left(E\left[y^i_{t+1}\right]\right)$ that

$$\begin{aligned} E\left[\hat{b}^i\left(y^i_{t+1}\right)\right] &\approx \hat{b}^i\left(E\left[y^i_{t+1}\right]\right) + \frac{1}{2}\hat{b}^i_{yy}\left(E\left[y^i_{t+1}\right]\right) Var\left[y^i_{t+1}\right] \\ Covar\left[\hat{b}^i\left(y^i_{t+1}\right), \hat{b}^j\left(y^j_{t+1}\right)\right] &\approx \hat{b}^i_y\left(E\left[y^i_{t+1}\right]\right)\hat{b}^j_y\left(E\left[y^j_{t+1}\right]\right) Covar\left[y^i_{t+1}, y^j_{t+1}\right] \end{aligned}$$

where $\hat{b}^i_y\left(E\left[y^i_{t+1}\right]\right)$ and $\hat{b}^i_{yy}\left(E\left[y^i_{t+1}\right]\right)$ denote the first and second partial derivatives of $\hat{b}^i(y)$ with respect to y when $y \equiv E\left[y^i_{t+1}\right]$. Now define the bond return vector $\boldsymbol{r_b}$ so that

$$r^i_b = \frac{\hat{b}^i\left(y^i_{t+1}\right)}{b^i\left(y^i_t\right)} - 1$$

It follows that[8]

$$E\left[r^i_b\right] \approx \left\{\frac{\hat{b}^i\left(E\left[y^i_{t+1}\right]\right)}{b^i\left(y^i_t\right)} - 1\right\} + \frac{1}{2}\frac{\hat{b}^i_{yy}\left(E\left[y^i_{t+1}\right]\right)}{b^i\left(y^i_t\right)} Var\left[y^i_{t+1}\right] \tag{8}$$

$$Covar\left[r^i_b, r^j_b\right] \approx \frac{\hat{b}^i_y\left(E\left[y^i_{t+1}\right]\right)}{b^i\left(y^i_t\right)}\frac{\hat{b}^j_y\left(E\left[y^j_{t+1}\right]\right)}{b^j\left(y^j_t\right)} Covar\left[y^i_{t+1}, y^j_{t+1}\right] \tag{9}$$

According to (8), the expected bond return is approximately equal to the return when the yield moves to $E\left[y^i_{t+1}\right]$ plus a correction term that is proportional to the convexity

[8] If for a bond k we have $T^k \leq t+1$, then $E\left[r^k_b\right] = \frac{100}{b^k\left(y^k_t\right)} - 1$ and, for all i, $Covar\left[r^i_b, r^k_b\right] = 0$.

of the bond, given by $\frac{\hat{b}^i_{yy}\left(E\left[y^i_{t+1}\right]\right)}{\hat{b}^i\left(E\left[y^i_{t+1}\right]\right)}$, and the variance of the yield. In particular, because bonds have positive convexity, this correction will always introduce a positive bias to the expected return.

Likewise, the covariance between the returns of two bonds is proportional to the modified duration of the bonds, given by $-\frac{\hat{b}^i_{y}\left(E\left[y^i_{t+1}\right]\right)}{\hat{b}^i\left(E\left[y^i_{t+1}\right]\right)}$, and the covariance of the yields. Of course, when $i = j$ we obtain the variance of each bond return.

Finally, since Theorem 1 gives a forecast of the entire yield curve at time $t+1$, $E\left[\boldsymbol{r_b}\right]$ and $Var\left[\boldsymbol{r_b}\right]$ can be readily computed and the forecast of the return of a bond portfolio is given by[9]

$$\begin{aligned} E\left[R_p\right] &= \boldsymbol{w}^T E\left[\boldsymbol{r_b}\right] \\ Var\left[R_p\right] &= \boldsymbol{w}^T Var\left[\boldsymbol{r_b}\right] \boldsymbol{w} \end{aligned}$$

where $\boldsymbol{w}_{m\times 1}$ gives the proportion of wealth invested in each bond and satisfy $\sum_{i=1}^{m} w_i = 1$.

Conclusions

Fixed income analysts deal constantly with the challenge of mapping their expectations about the general macroeconomic environment into movements of yield curves and ultimately into proper trading strategies. Given the complexity of this problem, many analysts prefer to develop first a forecasting model of a few benchmark yields and only then consider the problem of forecasting complete yield curves, if necessary.

This paper assumed that an analyst is able to provide forecasts of at least a few benchmark yields or combinations of yields. Then it constructed the yield curve that is consistent with the analyst's views, and computed the expected return of a bond portfolio in that scenario. Thus, the model proposed here is useful for a study of scenario analysis, when the analyst could estimate, for instance, the impact of expected developments in the macroeconomic environment on the yield curve and on the return of a bond portfolio.

The model builds on the theory of principal component analysis (PCA), can be easily extended to other markets and has no restrictions on the number of forecast variables or

[9]We assume that there are m bonds in the portfolio and that Theorem 1 is used to forecast the m corresponding yields-to-maturity. However, to do that we need to estimate the covariance matrix $\boldsymbol{S}_{m\times m}$ of yields from historical data. In practice, if the number of bonds is too large or if there is not enough historical data for all yields-to-maturity, it may be more appropriate to forecast the benchmark yields using Theorem 1 and use interpolation to compute the expected values and the covariance matrix of all yields.

the number of views. It also operates in the first two moments of the joint probability distribution of yields and makes no assumption about higher moments. This is an advantage relative to Bayesian theory, for instance, in which a parametric distribution is often assumed for the random variables.

One extension of the model could use ICA (independent component analysis) to derive the common factors driving yields (Hyvarinen et al., 2001). ICA works with independent factors (up to co-kurtosis) whilst PCA requires only that factors are uncorrelated. Thus, the forecast might work better with ICA when there is evidence of strong non-normality in the data. However, the benefit of applying ICA may be marginal given that the main source of error in the forecast arises from the analyst's views.[10] If the views are incorrect, there is little ICA could do to improve the forecast.

Appendix: Proof of Theorem 1

Define the normalised yield variations as $\Delta\tilde{\boldsymbol{y}} = \boldsymbol{D}^{-1}(\Delta\boldsymbol{y} - \boldsymbol{\mu})$ so that (3) can be re-written as

$$\boldsymbol{V}\boldsymbol{D}\,\Delta\tilde{\boldsymbol{y}} = (\Delta\boldsymbol{q} - \boldsymbol{V}\boldsymbol{\mu}) + \boldsymbol{\epsilon} \tag{10}$$

The principal components of normalised yield variations are defined as $\boldsymbol{p} = \boldsymbol{W}^T\Delta\tilde{\boldsymbol{y}}$. Then, using $\boldsymbol{W}^{-1} = \boldsymbol{W}^T$ we have $\Delta\tilde{\boldsymbol{y}} = \boldsymbol{W}\boldsymbol{p}$, which can be decomposed as

$$\begin{aligned} \Delta\tilde{\boldsymbol{y}} &= \tilde{\boldsymbol{W}}\tilde{\boldsymbol{p}} + \boldsymbol{u} \qquad (11) \\ \boldsymbol{u} &= \hat{\boldsymbol{W}}\hat{\boldsymbol{p}} \end{aligned}$$

where $\boldsymbol{p} \overset{def}{=} \left[\begin{array}{c} \tilde{\boldsymbol{p}}_{n\times 1} \\ \hline \hat{\boldsymbol{p}}_{(m-n)\times 1} \end{array}\right]$.

It is a standard result in PCA that $Var[\boldsymbol{p}] = \boldsymbol{\Lambda}$ and that a few principal components are enough to explain most of the variability of a collinear system. Thus, we assume that the future yield variations in (11) can be approximated by $\tilde{\boldsymbol{W}}\tilde{\boldsymbol{p}}$ and define $\boldsymbol{u}$ as the error in the approximation so that $E[\boldsymbol{u}] = \boldsymbol{0}$ and $Var[\boldsymbol{u}] = \hat{\boldsymbol{W}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{W}}^T$.[11]

Replacing (11) into (10) and re-arranging we have:[12]

$$\tilde{\boldsymbol{p}} = \left(\boldsymbol{V}\boldsymbol{D}\tilde{\boldsymbol{W}}\right)^{-1}(\Delta\boldsymbol{q} - \boldsymbol{V}\boldsymbol{\mu}) + \left(\boldsymbol{V}\boldsymbol{D}\tilde{\boldsymbol{W}}\right)^{-1}(\boldsymbol{\epsilon} - \boldsymbol{V}\boldsymbol{D}\boldsymbol{u}) \tag{12}$$

[10] Many thanks to Pete Eggleston, of the Royal Bank of Scotland, for raising my attention to this.

[11] As opposed to standard factor analysis, here the elements of $\boldsymbol{u}$ are not uncorrelated. See Jolliffe (2002, ch. 7) for a comparison of PCA and factor analysis.

[12] We note that $\boldsymbol{V}\boldsymbol{D}\tilde{\boldsymbol{W}}$ is a $n \times n$ square matrix and that it is invertible only if $rank(\boldsymbol{V}) = n$.

Under the assumption that $E[\boldsymbol{u}] = \boldsymbol{0}$, (12) gives a forecast for the first n principal components at time t. The first term in the right-hand side is the expected value of $\tilde{\boldsymbol{p}}$ and the second term is the error due to the analyst's uncertainty on the views and the approximation of $\Delta\tilde{\boldsymbol{y}}$ in (11) with the first n principal components.

It follows that $E[\boldsymbol{u}] = \boldsymbol{0} \Leftrightarrow E[\hat{\boldsymbol{p}}] = \boldsymbol{0}$ so that the forecast for the remaining $m - n$ principal components is zero. The rationale is that any yield curve movement implicit in the views is more likely to be caused by movements in the n most important principal components.

Finally, after replacing (12) into (11) and using $\boldsymbol{y}_{t+1} = \boldsymbol{y}_t + \boldsymbol{\mu} + \boldsymbol{D}\Delta\tilde{\boldsymbol{y}}$ the theorem follows under the assumption that $\boldsymbol{\epsilon}$ and $\boldsymbol{u}$ are independent. $\square$

References

Alexander, C. (2001). *Market Models: A Guide to Financial Data Analysis*. John Wiley & Sons.

Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43.

Diebold, F. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364.

Diebold, F., Li, C., and Yue, V. (2008). Global yield curve dynamics and interactions: A generalized nelson-siegel approach. *Journal of Econometrics*, to appear.

Hyvarinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. Wiley-Interscience.

Jolliffe, I. (2002). *Principal Component Analysis*. Springer, 2nd edition.

Loretan, M. (1997). Generating market risk scenarios using principal components analysis: methodological and practical considerations. In *The measurement of aggregate market risk*. Bank for International Settlements, CGFS Publications 7.

Meucci, A. (2005). *Risk and Asset Allocation*. Springer.

Nelson, C. and Siegel, A. (1987). Parsimonious modeling of yield curves. *Journal of Business*, 60:473–489.