1 Empirical Analysis of Term Structure Data

1.1 Introduction

Before looking at the empirical behaviour of yields, we need to introduce some notation. Let y_t^τ denote a set of yields that together form a yield curve, that is, a vector that stacks individual annual yields, with the same dating, t, but that are observed at different maturities, τ. In the practical examples included in this Element, we will typically use $\tau = \{3, 12, 24, \ldots, 120\}$ months, but τ can naturally take any value, at which yields are observed. When referring to a panel of yield observations (of dimension number of dates by number of maturities; i.e. a collection of yield curves observed at different dates), we will either write y, $y(\tau)$, or Y.

In a factor model, X will denote the extracted factors, and H, G, or B, will typically denote the matrix that translates factors into yields; this matrix is often denoted the 'loading' matrix because it expresses how each of the extracted factors impact, or load on, the yields at different maturities. Vector autoregressive models will be written as $z_t = m + \Phi \cdot (z_{t-1} - m) + e_t$, when written in mean-adjusted form, and sometimes as $z_t = c + \Phi \cdot z_{t-1} + e_t$, when written in constant form (i.e. $m = [I - \Phi]^{-1} \cdot c$).

At this point it may also be worth recalling that the yield curve is a by-product of the financial market trading process. Agents trade bonds that are quoted in prices, $p_t(\tau)$. A risk-free bond, the ones we primarily deal with here, guarantees to pay Eur 1 (in reality some scaling of 1, most often Eur 100) at the maturity of the bond. The price today is therefore, as always in finance, the discounted value of the promised payment that falls in the future: $P_t(\tau) = 1 \cdot (1 + y_t(\tau))^{-\tau} \Leftrightarrow y_t(\tau) = (P_t(\tau))^{1/-\tau} - 1$, in discrete time, and $P_t(\tau) = 1 \cdot e^{-y_t^\tau \cdot \tau} \Leftrightarrow y_t(\tau) = -\frac{1}{\tau} \cdot log\,(P_t(\tau))$, in continuous time.

We will model exclusively zero-coupon bonds. These bonds are important because they form the basis for fixed income pricing: since all coupon paying bonds can be expressed as portfolios of zero-coupon bonds (of relevant maturities), once we know the prices of zero-coupon bonds, we can also find the market-clearing price for all existing coupon paying bonds, assuming that there is no idiosyncratic risk attached to these bonds, such as, for example, illiquidity risk. Most often, however, we do not work with prices, but instead focus on rates/yields, that is, on the annualised percentage return the bond gives, if we hold it to maturity. As implied by its name, a zero-coupon bond does not pay any coupons during its life, and its cashflow stream is therefore simple, as illustrated in Figure 1 for zero-coupon bonds of one, two, and ten-year maturities.

Typically, we get zero-coupon data from Bloomberg, Reuters, and other data providers. These data are available at daily, weekly, and monthly

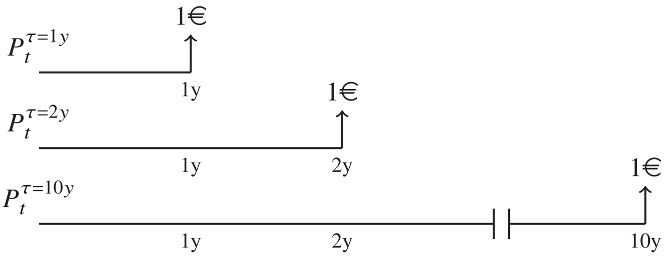

Figure 1 Zero-coupon cashflows

observation frequencies, and at predefined target maturities, for example, at $\{0.25, 1, 2, \ldots, 10, 15, 20, 30\}$ years.

1.2 Exploring Yield Curve Data

The example data used in this section are stored in the MATLAB workspace file named 'Data'. Data are obtained from public sources. The US data are downloaded from the Federal Reserve Board homepage.[1] These are the well-known and often-used Gurkaynak, Sack and Wright (2006) data. German yield curve data are obtained from the homepage of the German Bundesbank.[2]

For each segment, we have yields in per cent per annum across maturities, as well as model-based estimates for the expectations component and the term premium, both estimated at a ten-year maturity point. We will return to these latter two variables later on and for now only focus on the yield curve data. Let's load and plot these data: each data set contains monthly observations for the following variables: date and yields, and spans the period from January 1975 to December 2018, that is, a total of 528 time-series observations for each of the 6 included maturities per yield curve segment, which are $\{3, 12, 24, 60, 84, 120\}$ months.[3]

In addition to the time-series evolution of yields shown in Figure 2, it is also informative to see what the yield curve looks like in the cross-sectional dimension. For example, what does the average yield curve look like? And, what

1 https://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html
2 https://www.bundesbank.de/en/statistics/time-series-database
3 The shortest maturity observed for the raw German data downloaded from the homepage of the German Bundesbank is six months. For the illustrative examples shown in this section, it is simpler if observations at equal maturities are available for the German and the US yield curve data. Hence, an interpolation technique is used to obtain a three-month maturity observation for the German data. The exact process used is documented in the MATLAB code in Section 1. In an actual analysis we would not do this because it is not needed. Yield curve models can easily handle data observed at different maturities: we employ this trick here simply for expositional reasons.

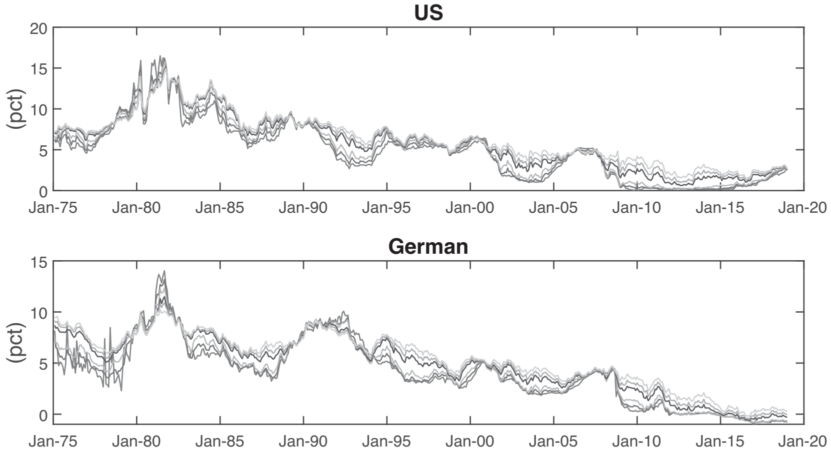

Figure 2 Yield curve data

The figure shows the time series of yields, observed monthly and covering the period from 1975 to end-2018, for maturities of six-month, one-year, two-year, five-year, seven-year, and ten-year for Germany, and for the US market the following maturities are shown three-month, one-year, two-year, five-year, seven-year, and ten-year. Yields for the USA, Germany, and the euro area are included in the plot. It is noted that the shortest maturity in the German market is six months (that is what is available from the German Bundesbank home page) while the shortest maturity available for the US data is three months.

are some of the most extreme shapes and locations that yield have displayed historically? These questions are explored in the following.

Note that one of the curves shown in Figures 3 and 4 may actually have materialised historically, since the calculations are done for each of the maturity points separately.

Going back to the time-series plots of the yields observed for the USA and German market segments, it is also interesting to observe that there is a very high degree of correlation among yields within a given market segment, and that a similarly high degree of correlation exists between market segments. It almost seems as if every little up- and down-ward movement in one maturity is mirrored by the other maturities in that market segment, with more pronounced movements the higher the maturity. Similarly, the secular swings that yields display over the twenty years of data are equally well visible across market segments.

A more structured view on the within and between segment correlation is illustrated in the following. For presentational purposes, correlations are shown only for a subset of the included maturities.

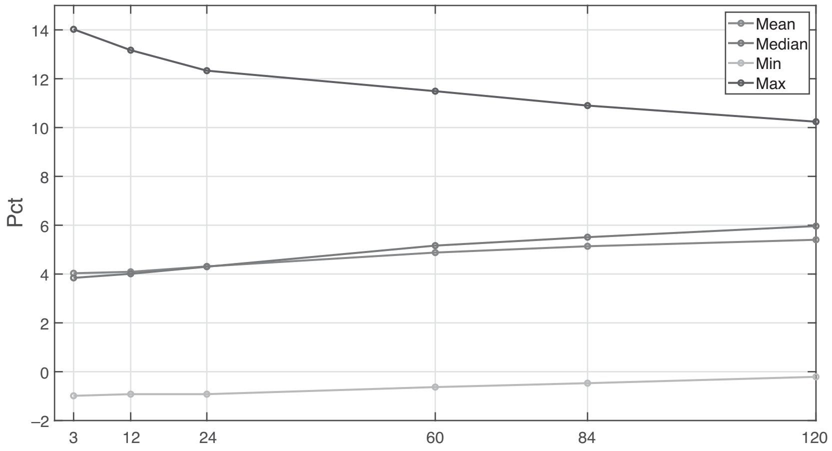

Figure 3 Generic shapes of the German yield curves

The figure shows the mean, median, min, and max of the German yields observed at a monthly frequency and covering the period from January 1975 to December 2018. The statistics are calculated across maturities. The x-axis shows the maturity of the yields in months (e.g. 120 months correspond to 10 years).

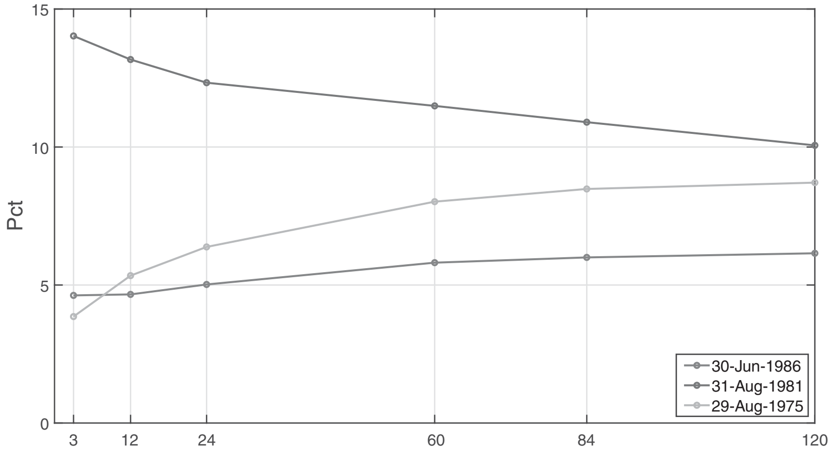

Figure 4 German yields with varying slopes

The figure shows German yield curves on the days when the slope ($y^{\tau=10y}$-$y^{\tau=3m}$) reached its minimum, maximum and average value, for the period from January 1975 to December 2018.

Figure 5 provides a visual representation of the correlation between German and US yields. If we had included other or additional yield curve segments, in addition to the three-month, five-year, and ten-year maturities, we would get qualitatively identical results. As expected based on the visual inspection of

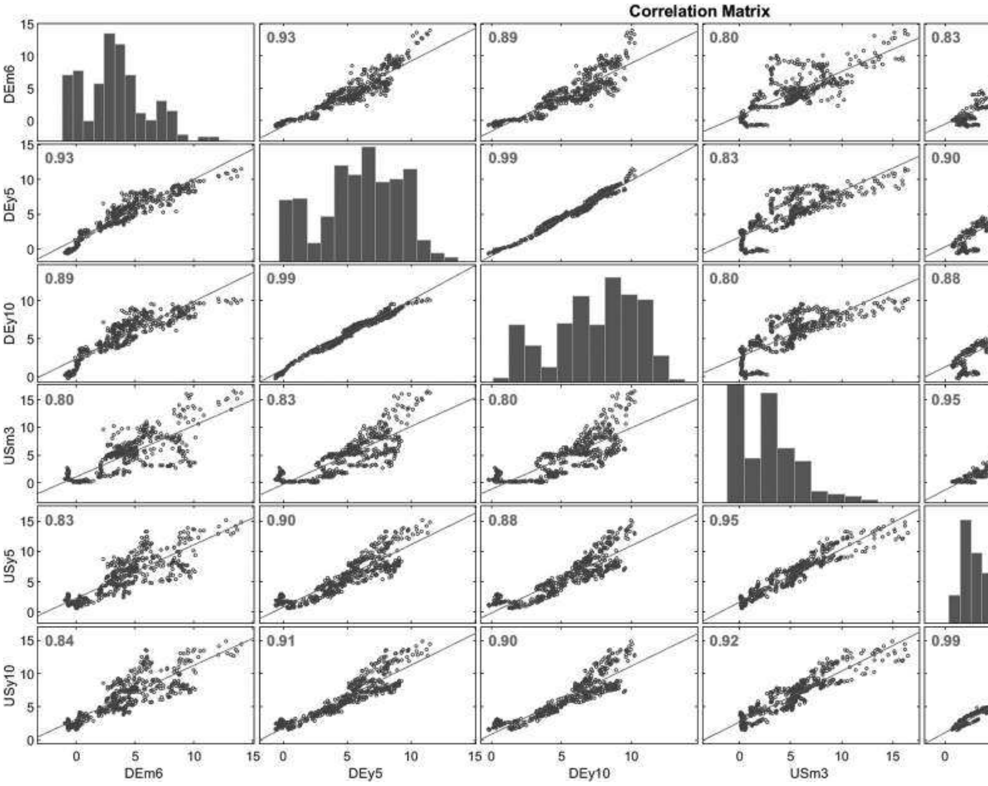

Figure 5 Correlations

The figure shows the pair-wise correlation between US and German yield levels observed at a monthly frec
1975 to December 2018. Correlations are calculated between the three-month, five-year, and ten-year maturity
red number indicates the correlation coefficient, and the red line shows the fitted regression line. On the diag

the time series plots, the cross correlations confirm our suspicion: yields within and across yield segments are very highly correlated. Note that a red number in the Figure 3 correlation matrix indicates that the correlation is statistically significant from zero at a 1 per cent significance level.

We could repeat the above correlation analysis for the first differences of the yield series – this would, for example, make sense, if yields were believed to be I(1) processes (i.e. integrated of order one). And, if we did this, we would obtain a correlation picture that is qualitatively identical to the one discussed here.

Now, looking at the time series plots of the yield curve segments, one could conclude, based on a preliminary and casual visual inspection, that the behaviour displayed by yields is somewhat different from what most people have in the back of their mind when they think about the trajectory of a stationary I(0) process. While this is a relevant thought, the discussion of stationarity will be taken up later on, when we discuss the eigenvalues of estimated vector autoregressive processes (VAR models – not to be confused with VaR, i.e. value-at-risk). For now, we treat observed yields as coming from a stationary data-generating process.

How can the overwhelming degree of correlation between yields be exploited? The answer is: by using principal component analysis (PCA)/ factor models. At this stage, it is worth noting that virtually all term structure models, as well as many other important financial models (e.g. ATP and CAPM for equity return modelling), rely heavily on PCA modelling principles. In fact, this econometric technique is quite possibly the single most important modelling idea in the field of quantitative time-series finance. To my mind, it is as important as PDEs (partial differential equations) are to the branch of finance that deals with derivative pricing. It is therefore fairly important to master this technique.

Before embarking on the factor modelling principle, it is worth spending a few minutes explaining why it is generally not advisable to use raw lagged yields directly to explain current yields. Doing this would amount to applying the following VAR-model set-up, where Y is a vector of yields, c is a constant, Φ is a matrix of autoregressive coefficients, and e is a vector of residuals:

$$Y_t = c + \Phi \cdot Y_{t-1} + e_t \tag{1.1}$$

Arguments against this modelling strategy are, amongst others:

- The number of yields modelled may vary from market to market and over time. It is therefore not clear which maturities that should be included in the model.

Cambridge University Press
978-1-108-97212-3 — A Practitioner's Guide to Discrete-Time Yield Curve Modelling
Ken Nyholm
Excerpt
More Information

- One may need to adapt the dimension of the model, depending on which market is modelled. This is inconvenient as well, as model results may not be comparable.
- Since correlation between yields is so high, we may run into the problem of multicollinearity.
- Projected yield curves and yield curve forecasts may turn out to violate standard regularities (e.g. individual yield curve points may be out of sync with the rest of the curve).
- The econometrician has very little control over the simulations; for example, it is difficult to steer the projections in a certain direction, if that is desired. Likewise, it is difficult to avoid certain (unrealistic) yield curve shapes and developments.

This last point is illustrated in Figure 6, using the German data.

It is dangerous these days to make statements about whether a given simulated yield curve has a realistic shape or not – and the future may prove me wrong – but despite what we have seen over the past years, I believe that the depicted simulated curves in Figure 7 are too oddly shaped to be considered for financial analysis (unless for some wild economic scenario): this applies to their shape and location, and to the overall simulated trajectory (Figure 6) for

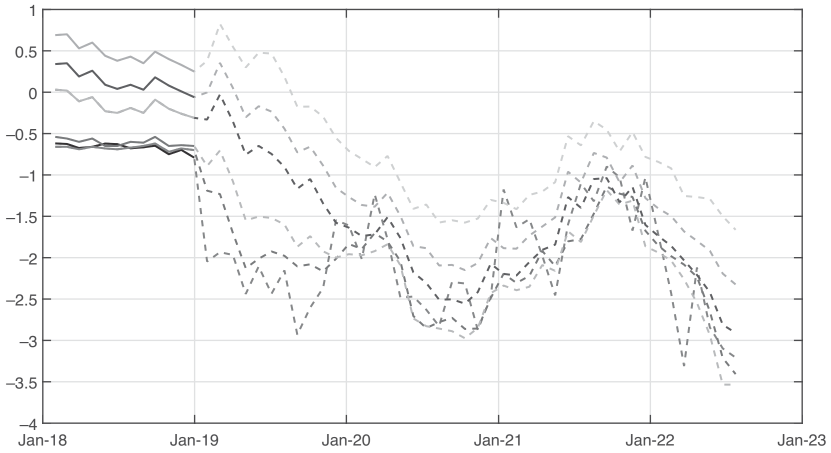

Figure 6 Naive yield curve forecasts

The figure shows how one can do naive forecasts of the yield curve, and what problems this may bring. A VAR model is fitted to individual maturity points using the full historical sample (from 1975 to end-2018) of German yields. Each maturity is then projected forty-two monthly periods ahead using the VAR. These projections are started at the last observation covered by the data sample.

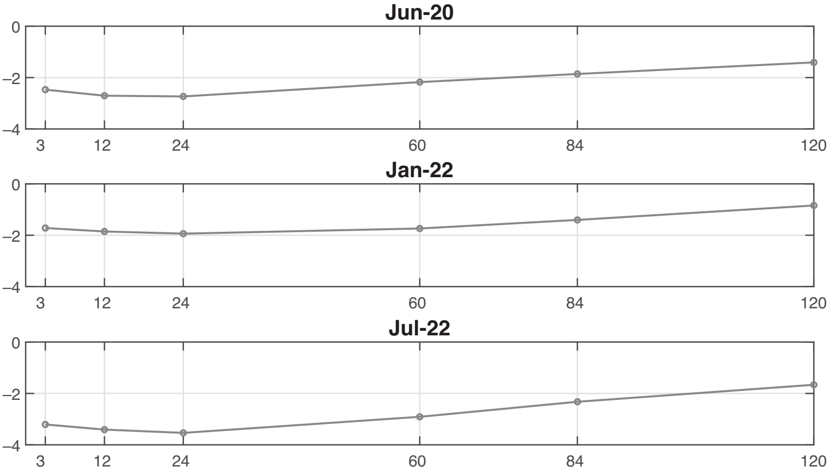

Figure 7 Randomly chosen projected yield curves

The figure shows randomly selected sample curves picked among the forty-two projected curves. The x-axis shows the maturities in months at which yields are recorded.

the yields over the coming forty-two months. One may of course have a rule-of-thumb and a routine that kicks out too oddly looking yield constellations and trajectories. Doing this will naturally change the distributional assumptions of the forecasted yields (since they are truncated at some pre-specified value) compared to the distribution exhibited by the historical data. One way to remedy this is to rely on approaches similar to Rebonato, Mahal, Joshi, Bucholz and Nyholm (2005), where historical residuals are block-bootstrapped (i.e. resampled) and the resulting simulated yields are smoothed to achieve shapes that are akin to those seen in historical data.

1.3 A First Look at Principal Component Models

Dimension reduction is one of the great feats of PCA / factors models: the core idea is that the majority of the variability of a given data set derives from a few underlying (sometimes not directly observable) factors. This concept is familiar; for example, the well-known CAPM prescribes that a single market factor is responsible for the expected return on all equities traded in the economy. Recall that the security market line is written as: $E[r_i] = r_f + \beta_i \cdot \left(r_m - r_f\right)$, where investors are rewarded only for taking market risk in excess of the risk free rate. r_m is the return on the market portfolio, that is, the underlying factor in this model, r_f is the observable risk free rate, and β_i is the sensitivity of the i'th security's return, r_i. In factor model language, r_f is the constant, r_m is the underlying factor, and β_i is the factor sensitivity that translates the factor observation into something that is applicable to the i'th security. We can naturally operate with more than one factor. Typically, term structure models include between one and five factors.

In general terms, and using matrix notation, we can write a factor model for the yield data in the following way:

$$\underbrace{Y_t}_{(\#\tau \times 1)} = \underbrace{G}_{(\#\tau \times \#F)} \cdot \underbrace{X_t}_{(\#F \times 1)} + \underbrace{\Sigma}_{(\#\tau \times \#\tau)} \cdot \underbrace{e_t}_{(\#\tau \times 1)} \quad e_t \sim N(0, I) \tag{1.2}$$

where Y is the vector of yields observed at time t, G is the aforementioned loading matrix, that translates the extracted factors X into into yields, Σ collects the standard deviations of the residuals, and $e\, N(0, 1)$. The dimensions of the variables are recorded below each entry, with the $\#$-sign referring to the dimension of each variable: so, for example, $\#\tau x1$ reads 'number tau by 1'. Say that yields are observed at the following $\{3, 12, 24, 36, 60, 120\}$ months: in this case we would have: $\tau\{3, 12, 24, 36, 60, 120\}'$ so, $\#\tau = 6$, and since this is a column vector, the number of columns included in τ is just 1. Consequently, the dimension of τ is therefore $6x1$. The dimension of the other included variables are denoted in a similar way, with $\#F$ representing the number of included factors. So, our first job when using factor models is to settle on an appropriate number of factors to extract (i.e. to choose $\#F$). We will always have that $\#F < \#\tau$, utilising the high cross-sectional correlation between yields, as shown in Figure 5, to reduce the dimensionlity of Y.[4]

Looking at the expression for Y_t in (1.2) indicates that if we know the factor loadings Φ, then we can find the factors X_t using linear regression, or by inversion. Underline the previous sentence! – we will use this 'trick' extensively when dealing with Nelson-Siegel type yield curve models later on. To preview a bit, let's quickly see how to back out the factors X using the full set of data – as mentioned, we will return to this issue in greater detail later on. First we write the above expression in terms of the full data set:

$$\underbrace{E[Y]}_{(\#\tau \times \#Obs)} = \underbrace{G}_{(\#\tau \times \#F)} \cdot \underbrace{X}_{(\#F \times \#Obs)}$$

where $nObs$ is the number of dates the data spans. Assume G is known. Then, in the context of an OLS regression, G represents the explanatory

[4] If a purely statistical factor model is estimated on the yield curve data, we will obtain factors that are orthogonal. However, later on it will become clear that the choice of G defines the economic interpretation that can be attached to the extracted yield curve factors; and in this case we will, in general, not find orthogonal factors.

variables and X the parameters to be estimated. We can therefore find X in the following ways:

$$\hat{X} = G^{-1} \cdot Y \tag{1.3}$$

or

$$\hat{X} = (G' \cdot G)^{-1} \cdot G' \cdot Y \tag{1.4}$$

where the first equation in (1.3) represents a pure inversion, and the second is the standard OLS formula. Returning to the main topic of this section (i.e. factor models), let's see if the DE and US data hide some interesting underlying patterns (i.e. factors), and let's try to construct a completely data-driven joint model for these to yield curve segments on the basis of such underlying factors.

The intention here is only to show how factor models can be useful for modelling term structure data, without infusing any term structure modelling knowledge – in other words, the illustrated strategy may be what an econometrician would choose to do if she had not received any term structure schooling. Later on in the Element it will become clear, that such an econometrician can actually be quite successful at modelling term structure data!

A clarification about the term 'factor models' is warranted here. When I refer to 'factor models' and 'factors', I do in fact mean 'principal Components', (i.e. the outcome of applying the PCA function in MATLAB). So, throughout, it is assumed that yield curve factors can be formed as a linear combination of observed yields. Alternatively, if a true factor modelling approach was applied, the starting point would be some underlying latent factors that were causing the evolution observed in the yield curve, and we would try to extract these factors. As we shall see, we will typically revert to factors that are directly interpretable in terms of yield curve observables (e.g. the level, slope and curvature of the yields curve), or actual maturity points on the yield curve. We will not, however, include unobservable quantities, such as, for example, the effective stance of monetary policy, or the natural long-term rate, as factors in the models that we work with in this Element.

Individual eigenvalues express how much of the overall variability in the data set the respective eigenvector explains. To help decide how many factors we need to include in our model, we can therefore link the number of factors to the overall variance that we want our model to capture.

Table 1 shows the cumulative fraction explained by the first six extracted principal components/factors explain of the US and German data. It is seen that four factors capture 100 per cent of the historical variability of both US and German yields, confirming the high degree of cross-sectional correlation among yield levels documented. If we believe that some of the variability in the observed data is due to noise, we should chose to model fewer than four