EUROPEAN CENTRAL BANK

EUROSYSTEM

Statistics Paper Series

Per Nymand-Andersen

Yield curve modelling and a conceptual framework for estimating yield curves: evidence from the European Central Bank's yield curves

No 27 / February 2018

Contents

Abstract

The European Central Bank (ECB), as part of its forward-looking strategy, needs high-quality financial market statistical indicators as a means to facilitate evidence-based and sound decision-making. Such indicators include timely market intelligence and information to gauge investors' expectations and reaction functions with regard to policy decisions. The main use of yield curve estimations from an ECB monetary policy perspective is to obtain a proper empirical representation of the term structure of interest rates for the euro area which can be interpreted in terms of market expectations of monetary policy, economic activity and inflation expectations over short-, medium- and long-term horizons. Yield curves therefore play a pivotal role in the monitoring of the term structure of interest rates in the euro area. In this context, the purpose of this paper is twofold: firstly, to pave the way for a conceptual framework with recommendations for selecting a high-quality government bond sample for yield curve estimations, where changes mainly reflect changes in the yields-to-maturity rather than in other attributes of the underlying debt securities and models; and secondly, to supplement the comprehensive – mainly theoretical – literature with the more empirical side of term structure estimations by applying statistical tests to select and produce representative yield curves for policymakers and market-makers.

JEL codes: G1, E4, E5

Keywords: term structure, yield curve models, data quality

1 Non-technical summary

It is important for central banks to have reliable representations of the term structure of interest rates available. One of the main advantages of this curve is that the underlying instruments are considered free of credit risk and therefore provide the floor for the cost of borrowing of the economy and serve as a useful means of measuring capital market interest rates.[1] The main use of the yield curve, from a monetary policy perspective, is to obtain an empirical representation of the term structure of interest rates, which can be interpreted in terms of market expectations of monetary policy, economic activity and inflation expectations over short-, medium- and long-term horizons. From an ECB monetary policy perspective, a yield curve enables the ECB to gauge the effect of information about asset prices on market expectations of future inflation and interest rates, and thus more transparently describe market developments. Furthermore, movements in prices of financial instruments affect economic wealth and economic sentiment and, via these channels, domestic spending decisions. Yield curves can also provide valuable information for other central banking purposes, in particular, as input for financial stability, financial integration, market operations analysis and banking supervision. Besides, once a nominal yield curve is computed, a term structure of real interest rates and break-even inflation rates can be derived. An important challenge when estimating yield curves is that they should, of course, reflect the relevant movements in the underlying term structure of interest rates, while, at the same time, not every possible data point can be captured in a smooth curve.

Against this background, this paper reviews the fundaments and two types of yield curve models available in the literature, and aims to draw out some lessons firstly for creating a conceptual framework for extracting high-quality bond samples and secondly for applying statistical tests for yield curve evaluations and selections. This paper confirms the importance of knowing the uniqueness of your dataset and focusing on data quality before starting any empirical testing. Data analysis and data quality is king in any scientific work. Data cleaning is too often neglected or de-prioritised in scientific work with the risk of not being able to differentiate noise from real (economic) signals when interpreting statistical results. The paper then develops a comprehensively empirical exercise for selecting an optimal yield curve for the euro area. The empirical evidence was gathered with the kind assistance of some euro area national central banks.

1 "Free of credit risk" refers to AAA-rated central government bonds. Yield spreads between euro area government bonds include a certain element of "credit risks". Nevertheless, the German (Bund) curve has become a widely accepted proxy for risk-free yields in the euro area (see the July 2014 issue of the ECB's Monthly Bulletin).

Four types of yield curve models have been tested; two spline-based and two parsimonious models:

Yield curve modelling

Spline-based models[2]

- The **Waggoner** cubic spline method with a three tiered step-wise linear penalty function
- The **variable roughness penalty** (VRP) method, which is based on the spline-based technique proposed by Waggoner but with a continuous penalty function. It can therefore also be referred to as a Waggoner model with a smooth penalty function

Parametric (parsimonious) models[3]

- The Nelson & Siegel model
- The Svensson model

All four models were tested using the same dataset covering two years of daily euro area government bond prices and yields, split into an in-sample and an out-of-sample dataset. All models were tested not only for the full two years of daily data, but also for more limited time periods under specific market conditions[4], and were assessed at different maturity spectrums.

Three statistical tests and criteria were applied to evaluate the performance of the models:

Smoothness test

- This statistical test is done in order to ensure the best overall data fit without trying to fit every data point.

---

2 These types of models have been used by, among others, the Bank of England, the Federal Reserve Bank of New York, the Bank of Japan and the Bank of Canada.

3 These types of models have been used by, among others, the Deutsche Bundesbank, the Banco de España, the Banca d'Italia and the Banque de France.

4 The periods were: (1) periods of ECB interest rate cut expectations; (2) periods of ECB interest rate increase expectations; (3) periods of high market volatility; and (4) periods of widening credit spreads.

Flexibility & goodness-of-fit test

- These statistical checks are performed to test if the model estimations adequately capture the movements in the underlying term structure, in particular for shorter maturities.

Robustness test

- This statistical test is done in order to ensure that changes in the data at one maturity do not have a disproportionate effect on the fit for other maturities.

With regard to the bond sample population, the empirical exercise confirmed that it is vital for practitioners to complete descriptive statistics and to obtain a comprehensive overview and understanding of the dataset before conducting any empirical studies. Bond types, prices and liquidity considerations play a major role in the cleaning process, as the more liquid a market is, and the more frequent the trades, the more representative the informational content of prices in the market is. Furthermore, the volatility of yields across classes of residual maturity was analysed, which led to the elimination of bonds with a residual maturity of less than three months, as this class showed significantly higher volatility than other maturity classes, combined with the fact that the information content of bond yields close to maturity is reduced. Descriptive statistics of the characteristics of the single bonds were performed to eliminate unexpected values and bonds whose yields represent abnormal levels and deviate by more than twice the standard deviation of the yields of similar bonds in the same maturity brackets. Finally, the descriptive statistics assisted in the provision of a conceptual framework with fundamental recommendations on how to derive a clean and high-quality dataset for yield curve estimation purposes.

Applying the model estimations using the high-quality bond and statistical tests demonstrated that all four models are able to reliably and consistently represent the term structure of interest rates over the reference period, with low error statistics.

Concerning the parsimonious models, the Svensson model performed slightly better than the Nelson & Siegel model in terms of the flexibility & goodness-of-fit test for both the in-sample and the out-of-sample tests. This confirms similar results in the literature. Similarly, both parametric models yield quite smoothly sloping curves and showed very stable and similar test results for the smoothness and robustness of the curve. The spreads between the maturities became somewhat more volatile if yields on bonds with the shortest (three months) and longest (30 years) maturities were included. All in all, the empirical study concludes that the Svensson model performs slightly better than that of Nelson & Siegel.

Both spline-based models (Waggoner and VRP) provided results that are very similar and consistent with the parametric models in terms of goodness of fit, though with slightly lower error statistics and higher “hit rates” across maturities up to ten years. The slopes of the curves were also rather smooth for short- to medium-term spreads, but became more volatile when long-term government bond yields were included. The statistical tests also demonstrated a significant drop in hit rates for maturities above ten years. Over the whole two-year period, the Waggoner model seems to produce a slightly lower (though insignificant) fit than the VRP for maturities up to ten years, although the available evidence is not conclusive and depends on the statistical tests applied and the maturity band. The complexity of the optimisation process and the high number of parameters involved in the spline-based methods might imply a lower degree of transparency of the spline-based methods compared with the parametric ones which are more complicated to interpret, communicate and present to policymakers and the general public.

The parsimonious models and their results are more transparent and the parameters can be more easily interpreted, which serves the purposes of communication and accountability in releasing yield curves to the general public. The ECB therefore releases two credit risk yield curves using the Svensson model each day at 12:00 CET. The spot, forward and par yield curves, as well as their corresponding parameters, are released as time series for each curve. The two curves are estimated using a high-quality dataset for government bond prices and yields reflecting different credit default risks, applying the conceptual framework proposed in this paper.

The daily yield curves are released daily on the following ECB webpage: http://www.ecb.int/stats/money/yc/html/index.en.html

2 Introduction

Financial markets are a key channel for the transmission of monetary policy impulses to the real economy, and the changes in financial markets can reflect agents' expectations about future macroeconomic developments. Therefore, financial market indicators are intensively used in order to systematically analyse the relationship between monetary policy and the euro area financial markets' structure and dynamics, and play a contributory role in the ECB's monetary policy decision-making process and strategy. One could argue that the quality of monetary policy decision-making is dependent on, among other things, the availability and quality of financial market indicators and their explanatory power. Therefore, a broad range of euro area economic and financial market indicators are made available and analysed.

This paper focuses on one such financial market indicator, namely the calculation of euro area government bond yield curves tailored to the needs of the European System of Central Banks (ESCB). Furthermore, the paper contributes to the transparency of the statistical ingredients of calculating and releasing daily yield curves and thus provides a source of daily statistics to private and public financial agents and the general public.

The ECB needs reliable representations of the term structure of interest rates. One of the main advantages of a single representative curve for triple A-related euro area government bonds is that the underlying instruments are considered to be almost free of credit risk. Therefore, such a curve provides a floor for the borrowing costs in the economy and provides a useful benchmark for assessing market interest rates. Besides, yield curves can be used to gauge market expectations concerning monetary policy, economic activity and inflation over short-, medium- and long-term horizons. Yield curves can also provide valuable information for other central banking purposes, in particular, as input for financial stability, systemic risks, financial integration and market operations analysis. In addition, once a nominal yield curve is computed, a term structure of real interest rates and break-even inflation rates can be derived and regularly released. It is therefore important to estimate a yield curve where changes mainly reflect changes in the yields-to-maturity rather than in other attributes of the underlying debt securities. In particular, there must be enough observations available to estimate the curve with a sufficient degree of precision and it must not be affected by changes in perceived credit risk.

Against this background, this paper presents the rationale for the ECB to release daily euro area yield curves based on the Svensson model, while elaborating on the importance of performing substantial descriptive statistics on the bond sample and enhancing the data quality prior to yield curve estimations.

Chapter 3 provides a synopsis of the main yield curve theories and models, including a comparison of their advantages and weaknesses. Chapter 4 provides an operational framework for obtaining a high-quality bond sample and evaluates the statistical properties for estimating yield curves. Chapter 5 presents the empirical evidence and compares the performance of the four models. Finally, Chapter 6 concludes the empirical work by selecting one model for estimating euro area yield curves on a daily basis.

3 Yield curve methods for central banking purposes

Euro area financial market indicators are valuable for monetary policymakers. Movements in prices and volumes of financial instruments affect economic wealth and economic sentiment and, via these channels, domestic spending decisions. Financial instrument price movements are also indicative of changes in the expectations of the private sector regarding economic prospects. More specifically, yield curves are estimated to provide a graphic representation of the relationship between the returns and the terms-to-maturity of debt securities at any given time. The information content of a yield curve reflects the asset pricing process on financial markets. When buying and selling bonds, investors include their expectations of future inflation and real interest rates and an assessment of risks. An investor calculates the price of a bond by discounting the expected future cash flows. Usually, the term “yield curve” refers to the term structure of interest rates of zero-coupon bonds without default risk.

The yield curve offers a useful set of information for monetary policy purposes and gauges information about the expected path of future short-term rates and the outlook for economic activity and inflation. The relative level of short- and long-term interest rates at a certain date depends on market participants’ expectations of future short-term interest rates. Therefore, the slope of the yield curve has often appeared to be a useful indicator for predicting future economic activity. A steepening of the curve often anticipated an acceleration of economic activity while a flattening, and in particular an inversion, of the curve often indicated an imminent slowdown. The explanation is that a large positive spread between long- and short-term interest rates may indicate that the market anticipates an increase in short-term interest rates because of a more positive outlook for economic growth. In addition to growth expectations, the longer end of the yield curve may also mirror market participants’ views about trend developments in inflation.

The theories underlying the term structure of interest rates can be briefly summarised as follows:

Liquidity preference theory – this theory indicates that investors are risk-averse and will demand a premium for holding securities with longer maturities. This relates to the principle that investors prefer to have cash available today rather than tying up cash for a future deliverable and in compensation therefore require a risk premium. All other things being equal, one would therefore expect to see a rising slope of the yield curve.

The pure expectations hypothesis – according to this theory the forward rates govern the curve; these are simply expectations of future spot rates and do not take into account risk premia. In other words, the hypothesis indicates that long-term interest rates can act as a predictor of future short-term interest rates. Instead of buying a long-term bond, an investor could also consider rolling over investments in short-term bonds over a period of the same length as the remaining maturity of the long-term bond. Disregarding risk considerations, the total return on the investment in the long-term bond should be equal to the expected cumulative return on the revolving investment in short-term bonds. This also implies that the average expected future short-term interest rate over the investment horizon should equal the long-term interest rate. For example, an upward-sloping yield curve, featuring higher long-term interest rates than short-term interest rates, would then imply an expected increase in short-term rates. Empirical evidence suggests this hypothesis often overstates future short-term interest rates, which may be caused by the remaining risk premium. Hence, in order to extract market expectations about future short-term interest rates from the yield curve, these risk premia need to be estimated. However, this task is complicated, in particular, by the fact that risk premia seem to fluctuate over time. Nevertheless, the expectations hypothesis still appears to be a reasonable starting point for gauging interest rate expectations from the yield curve. Notably, for shorter horizons, risk premia tend to be relatively low and stable during normal market conditions.

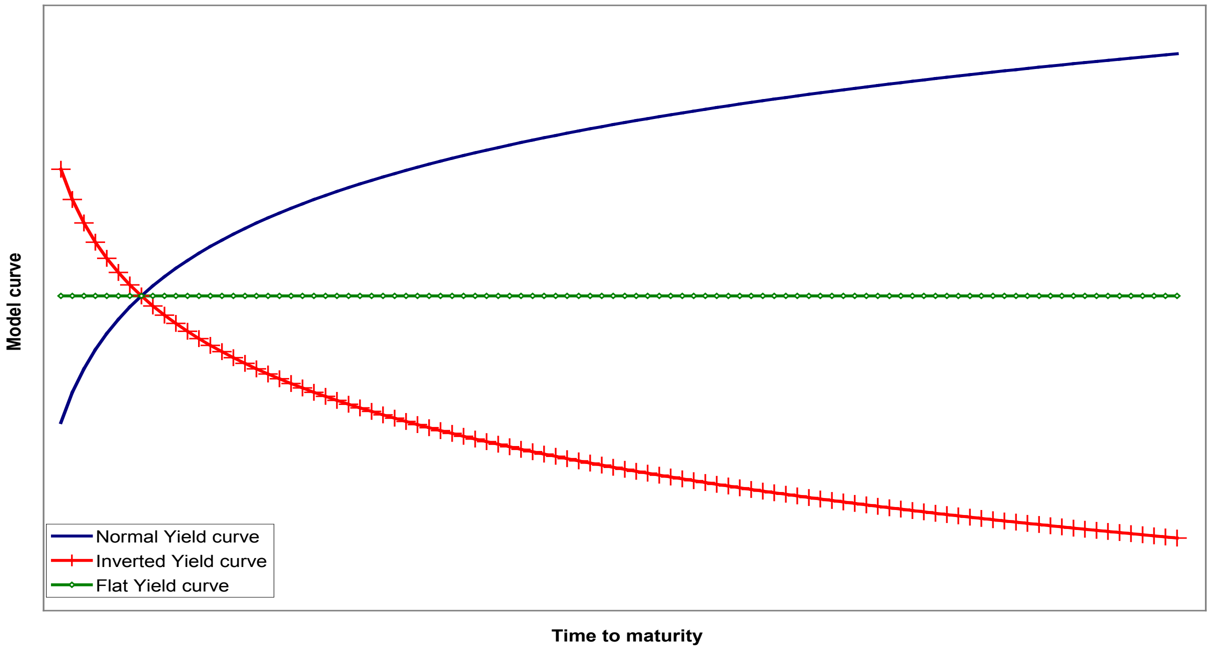
Segmented markets hypothesis – the yield curve depends on supply and demand in different sectors and each sector of the yield curve is only loosely connected with the others. This hypothesis is based on the belief that each bond market segment is largely populated by investors with a particular preference for investing in securities within that maturity time frame – short-term, intermediate-term or long-term. The yield curve is therefore shaped by the factors of supply and demand at each maturity length.

Preferred habitat theory – according to this theory, which is related to the segmented markets hypothesis, investors may also have a maturity preference, and will shift to another maturity if the increase in yield is deemed sufficient to compensate for the shift.

The yield curve depicts interest rates with different remaining maturities as shown in Chart 1. The yield curve shifts and changes in shape as a direct response to market movements (shocks), which can have level, slope and curvature effects on the curve. The level effect describes how the interest rate changes by the same amount for all maturity levels, whereas slope changes reflect the case where the short-term or long-term rate changes are relatively disproportionate to each other. The main effect of curvature relates to the medium-term interest rates, where the yield curve becomes more hump-shaped than previously.

Chart 1

Different shapes of the yield curve



In normal circumstances, yield curves are upward-sloping, reflecting the fact that longer-term securities give a higher rate of return than shorter-term securities, since the longer the lender has to wait for the repayment of his/her loan, the higher the expected risk (or term) premium.

Assuming a risk-free bond with known maturity and coupon payments and that the interest rates at the different times are known, then the price of a coupon-paying bond can be written as shown below.

$$P_t = \sum_{m-1}^{M} \frac{C}{(1 + s_{t,m})^m} + \frac{N}{(1 + s_{t,M})^M}$$

Where: C = Coupon

N = Redemption value

S = Interest rates

M = Maturity

In this case, the credit risk is negligible and the interest rate is solely determined by expectations of inflation and real interest rates. If held to maturity, the discount rate equals the yield-to-maturity (or the internal rate of return). The yield-to-maturity of a bond is the discount rate that would make the sum of the present values of all the future cash flows equal to the bond price. The computation can be made on the basis of market prices for the underlying financial instruments, for instance government bonds that are traded on stock exchanges and other trading platforms. Calculating this interest rate is straightforward in the case of zero-coupon bonds, which provide only one payment. However, in practice, limited numbers of zero-coupon bonds are available within the euro area, especially for residual maturities exceeding 12 months. Most euro area government debt securities are coupon-bearing. When calculating the yield-to-maturity of coupon-bearing bonds, all payment flows (coupons and redemptions) are discounted to current values at the same rate – i.e. the yield-to-maturity. In order to derive the implicit average annual interest rate from the market price of a coupon-bearing bond, each future interest payment on this bond must be discounted by the different current average interest rates related to the time at which the future payment occurs. This entails solving a set of equations with several unknowns. To facilitate the term structure estimation, it is useful to impose a functional form between interest rates and time to maturity. The term structure is then found via an iterative procedure.

A number of estimation methodologies exist, some more complex than others, with which the zero-coupon and forward rate curves from observed (bond price) data can be derived. Anderson et al. (1996) categorise these into two distinct groups: first, models that make specific assumptions about changes in state variables and asset pricing methods using either equilibrium or arbitrage arguments; second, models based on statistical techniques where the current yield curve is described by "smoothing" the data obtained from asset prices[5]. However, each technique, whether from the former or latter group, can provide surprisingly different shapes for these curves. As a result, estimation technique selection depends primarily on its final use.

Anderson et al. (1996) and Bolder and Gusba (2002) present the different approaches as:

1. the McCulloch (1971, 1975) method, which fits a cubic spline to the discount function using an implicit smoothness penalty;

2. the Fisher, Nychka and Zervos (1995) (FNZ) method, which fits a cubic spline to the forward rate function and makes the smoothness penalty explicit by imposing a (constant) roughness penalty;

3. the Waggoner (1997) approach, which differs from FNZ in that it introduces a variable roughness penalty;

4. Anderson and Sleath (2001) create a variant of the Waggoner model with a different roughness penalty functional form. Anderson and Sleath chose a more complex penalty function which varies continuously over maturities and in which only three parameters need to be estimated, compared with Waggoner's five smoothing parameters;

5. the parametric approach put forward by Nelson & Siegel (1987) and extended by Svensson (1994), which fits an exponential approximation of the discount rate function directly to bond prices.

5 i.e. removing the noise from the data.

The first four methodologies are spline-based techniques[6], whereas the fifth approach is a parsimonious parametric[7] approach. The different approaches will, in essence, involve a trade-off between flexibility to accommodate genuine bends in the term structure and "smoothness". The two approaches are summarised and compared in the next three sections.

3.1 Spline-based models

Spline-based methods model a curve using a piece-wise cubic polynomial, with segments joined at so-called knot points (McCulloch, 1971). Further developments in these techniques apply constraints to ensure that the pieces join up and look smooth (FNZ, 1995; Waggoner, 1997). Spline-based models were pioneered by McCulloch (1971, 1975). The method put forward by McCulloch is one in which the discount function is estimated using a regression spline.

The significant step forward made by this method is the concept of a basis spline. Computing and estimating piece-wise polynomials allow the individual curve segments to move almost independently of each other (subject to the continuity constraints) so that separate regions of the curve are less affected by movements in nearby areas.

Fisher, Nychka and Zervos (1995) (FNZ) developed a technique that fits smoothing splines instead of regression splines that include a penalty for excess "roughness" to extract the forward rate curve. This roughness penalty is constant across maturities, and, as a result, the spline is stiffened, which in turn reduces oscillatory behaviour. The recommended number of nodes is approximately one-third of the number of bonds used in the estimation, and the nodes should be spaced so that roughly an equal number of bonds mature between adjacent nodes. Since the penalty forces an implicit relationship between the parameters of the spline, the actual number of parameters is reduced. However, the FNZ method also reduces the fit, and Bliss (1997) found that the use of a constant roughness penalty tends to misprice short-term securities. He argues that this does not allow for greater flexibility, which is necessary at the short end, where there is more true curvature in the term structure.

Waggoner (1997) follows a similar approach to that taken by FNZ in that he uses a cubic spline to approximate the forward rate function. His approach, though, differs in that instead of assuming that the smoothing penalty is invariant to maturity but variable over time, he uses a penalty that varies over maturities, and chooses a three-tiered step-wise function for his smoothing parameter, with steps at one and ten years to maturity. This approach thus dampens the oscillations at the long end,

[6] Anderson et al. give an intuitive explanation of spline: "… a polynomial spline can be thought of as a number of separate polynomial functions, joined 'smoothly' at a number of join, break or knot points" (Anderson et al. (1996), "Estimating and Interpreting the Yield Curve", page 25).

[7] A parametrically parsimonious model for yield curves is able to represent the shapes generally associated with yield curves: monotonic, humped and S-shaped (Nelson & Siegel, 1987).

whilst retaining flexibility at the short end. Moreover, he uses the same approximate number and spacing of node points as FNZ.

Anderson and Sleath (2001) create a variant of Waggoner's model with a different roughness penalty functional form. Waggoner, taking into account the US government bond market structure, chose a step-wise penalty function, which penalises the short-, mid- and long-term maturities with the respective constant penalties. Waggoner has five smoothing parameters to estimate (two for the maturities at which there is a step, and three for the step levels). Anderson and Sleath chose a more complex penalty function, which varies continuously over maturities, but in which only three parameters need to be estimated.

Vasicek and Fong (1982) also use a spline-based model to derive the discount function. They modify the exponential form of the discount function to fit bond price data, from which yields are derived. Mastronikola (1991) fits a par yield curve, but the drawback of such a model is that it does not allow for bonds with the same maturity date to be discounted at the same rate. A possible drawback of these models is that they do not seem to overcome the discrepancy between the theoretical mean yield curve and the observed curves; the latter are substantially more concave than implied by the theory. This could be solved by using multifactor affine models, which consider bond yields as a function of several macroeconomic and financial variables (Backus et al., 1998, Campbell et al., 1997, or Cassola and Barros, 2001). Brousseau (2002) demonstrates, however, with the Duffie and Kan model (one-factor model with five parameters) that the actual observed yield curve fits the theoretical pattern with a precision of only a few basis points. Brousseau argues that this precision could also be explained by factors other than the model itself.

3.2 Parsimonious functional forms

The N&S model is a parametric model which specifies a functional form for the instantaneous forward rate, f(t), as follows:

$$f(t) = \beta_0 + \beta_1 exp\left(-\frac{t}{\tau}\right) + \beta_2 \frac{t}{\tau} exp\left(-\frac{t}{\tau}\right)$$

This model is able to capture the stylised facts, describing the behaviour of the forward rate curve. The parameters of this model can be interpreted as follows; β_0 is the long-term asymptotic value of f(t) for the interval of estimation and must be positive. β_1 is the spread between the long and short term and hence $\beta_0 + \beta_1$ is equal to the short-term rate (the rate at zero maturity). Furthermore, τ_1 specifies the position of the first hump or U-shape. β_2 determines the magnitude and direction of the hump.

The original motivation for this model was to create a parsimonious model of the forward interest rate curve that could capture the range of shapes generally seen in yield curves: monotonic form and with humps in various areas of the curve. This method allows the yield curve to be described by a few parameters representing the long-run level of interest rates, the slope of the curve and humps in the curve. The N&S and Svensson extension is basically the McCulloch model constrained to prevent the forward curve from taking undesirable shapes.

An important property of the model is that the forward rate asymptotes horizontally at the long end, because the expected future interest rates in 20 to 25 years are assumed to be indistinguishable. This methodology for estimating the forward curve was found to be sufficient to give a close fit to the data since real yields tend to converge to a constant level at relatively long maturities. In effect, this means that the real yield curve is flat over all but the shortest maturities. This real forward curve then translates into a real discount function.

To increase the flexibility and improve the fit of the N&S model, Svensson added a fourth term.

$$f(t) = \beta_0 + \beta_1 exp\left(-\frac{t}{\tau_1}\right) + \beta_2 \frac{t}{\tau_1} exp\left(-\frac{t}{\tau_1}\right) + \beta_3 \frac{t}{\tau_2} exp\left(-\frac{t}{\tau_2}\right)$$

The model has two more parameters than the Nelson & Siegel model. As β3 is analogous to β2 (as expressed above for the N&S model), the additional parameters can be interpreted as determining the magnitude and direction of the second hump. τ2 specifies the position of the second hump or U-shape.

3.3 Comparing spline-based models with parsimonious models

Waggoner tested the McCulloch, FNZ and Waggoner's VRP methods by comparing their in-sample and out-of-sample performance in pricing bonds. He found that the FNZ method performs slightly better than that of McCulloch in pricing securities with more than one year to maturity; however, the opposite is true for pricing securities with less than one year to maturity.

Waggoner finds that flexibility[8] is retained at the short end of the curve whilst it has a dampening effect on the oscillation at the long end, thus helping to make this methodology perform better than the McCulloch method across all maturities and better than the FNZ method across short maturities. Waggoner concludes, however, by pointing out that the results produced by the McCulloch method and the roughness penalty method (whether variable or constant) are very similar.

Anderson et al. (1996), in their comparative summary, take four different methods for working out a zero-coupon yield for the United Kingdom. They find that the resulting four shapes of the estimated yield curve do not seem to differ very much. However, the forward rate curves of each model are quite different. In both the N&S and Svensson models, for instance, the implication is that the forward rates gravitate smoothly towards a flattened long end. In the McCulloch model, on the other hand, the forward rates fluctuate according to maturity, ascending steeply as the term-to-

8 This flexibility is driven by Waggoner's use of a variable roughness penalty.

maturity lengthens. Finally, an examination of the forward rate curve in the Mastronikola (1991) model shows that it flattens at the long end, simultaneously exhibiting greater curvature than the N&S and Svensson models.

Anderson and Sleath (1999) provide a comparative summary of the techniques mentioned above. In particular, they compare the N&S and Svensson methods with the spline-based method put forward by FNZ and Waggoner. They raise the question whether the N&S and Svensson methods, although they provide smooth yield curve estimates due to their parsimonious nature, are sufficiently flexible to capture movements in the underlying term structure. They find that the Svensson model, due to the imposition of an extra parameter, is superior in the out-of-sample performance to the N&S model. In comparing the fit of the two spline-based techniques, Anderson and Sleath find that the variable roughness penalty curve proposed by Waggoner outperforms the FNZ curve. Intuitively, they conclude that this is because the FNZ suffers from the same lack of flexibility at the short end as the N&S method. Moreover, they find that the main differences between the variable roughness penalty and the Svensson model relate to the robustness criterion and constraints imposed at the long end. The Svensson model is constrained to converge to a constant at long maturities. The variable roughness curve, in contrast, is constrained only to be very smooth at these maturities. Anderson and Sleath further show that the Svensson curve changes dramatically, particularly at the short end when a single data point at the long end is changed. Curves based on parametric models are generally less well-suited to identifying abnormalities in individual maturity segments or individual bonds.[9]

Bliss (1996) tests and compares – in sample and out of sample – five distinct methods for estimating the term structure: the unsmoothed and smoothed Fama-Bliss method, the McCulloch method, the FNZ method and the Svensson method. He finds that the FNZ, both in sample and out of sample, performs badly compared with the other methods. He highlights that the FNZ method has systematic problems in handling short maturities and is susceptible to measurement errors in the data. He concludes that the parsimonious smoothed Fama-Bliss and Svensson method, as well as the less parsimonious McCulloch cubic spline method, performed comparably to each other.

Generally speaking, as the comparative studies above have shown, there is a continuous trade-off between smoothness and the fit factor in yield curve estimates, depending on which model is used. It is necessary, then, to strike a balance between those models that, on the one hand, are too flexible, over-fit the data or take outliers as the norm, and those that, on the other, are too parsimonious. As part of any selection process, the trade-off between smoothness and the fit factor in yield curve estimates depends on the intended final use of the yield curve.

A yield curve, estimated for monetary policy purposes by applying the aforementioned techniques, should not fit every possible data point (and outlier), but should represent a smooth curve as an indicator of market expectations.

[9] See Schich (1997).

Nevertheless, the curve should reflect actual expectations and therefore be flexible enough to capture movements in the underlying term structure.

Within the European Union, the national central banks of several Member States estimate yield curves and, in some cases, these national yield curves are also released to the general public.[10] It can be seen that both spline-based and parsimonious models are used by several national central banks and, within the euro area, the Nelson & Siegel and Svensson models are dominant. The instruments selected are mainly money market instruments and government bonds and, in some cases, these instruments are combined and used within the same curve. The yield curves have mainly been produced since the beginning of the 1990s and, in one case, data availability go back to 1972.

The diversity of the methodologies applied by central banks, in combination with the significant expertise and knowledge available within the ESCB, was one of the reasons for conducting the empirical exercise as part of developing a single conceptual and consistent approach to estimating yield curves that represent the euro area as a whole. In the post-crisis environment, the demand for national yield curves (complementing euro area yield curves) and the desire to explore the behaviour at near zero nominal interest rates[11] have significantly increased, although these are outside the scope of this paper.

[10] For an overview of national practices, please refer to BIS (2005), "Zero coupon yield curves: technical documentation", BIS Papers No 25. Other European authorities, such as EIOPA (the European Insurance and Occupational Pensions Authority) are calculating risk-free interest rate term structures for insurance obligations.

[11] See Krippner (2015).

4 Conceptual framework for testing yield curves

In this chapter, the focus is on completing descriptive statistics as part of getting to know your dataset. These descriptive statistics help to understand the characteristics of the datasets and the patterns and to identify abnormalities, which can provide guidance for removing observations from the dataset prior to starting the empirical exercise. It is a necessary part of the process of preparing a high-quality dataset that is fit for purpose. A few illustrative examples are included to facilitate the description of the statistics and the selection process.

4.1 Descriptive statistics and knowing your bond sample

In the absence of significant literature on data sampling and empirical considerations for selecting yield curve models, the following chapter provides several methodological recommendations for conducting, comparing and selecting a meaningful bond sample which can be used for estimating yield curves. These recommendations will be supported by data examples. Within the literature, the methodology and treatment of data selection is often neglected and there are risks that the conclusions of model testing will be open to question, as the results may be more biased by actual changes in the dataset or individual bonds than by differences between the results generated by the models. This is important, as financial market indicators provide leading and supplementary information in any decision-making process. Therefore, nothing is more important for the reliability and predictability of financial market indicators than using good quality data supported by descriptive statistics. This is particularly challenging for developing and providing “euro area” financial market indicators, as the euro area is one of the largest economic areas in the world, and releasing euro area government bond yield curves therefore represents more than the national contributions added together. This chapter will review statistical methods which can be applied for selecting a reliable and consistent euro area bond population. After that, the chapter will present a set of comparable evaluation criteria which can subsequently be used for testing yield curve models. Throughout this paper, the conceptual framework is presented with practical examples so as to support the reader in developing a deeper understanding of yield curve estimations. These practical examples were developed with the kind assistance of the ESCB Task Force on Financial Market Statistics, composed of representatives from the central banking community[12]. The dataset is used as an illustration of the empirical work and is therefore time independent. Nevertheless, during the period when the Monetary Union was established and until the financial crisis in 2007, the yield spreads between euro area government bonds experienced significant periods of continuous convergence. Following the financial crisis, the euro

[12] The contributing members are mentioned in the “Acknowledgements” section.

area experienced a significant widening of yield spreads among euro area government bonds. This has led to different and larger statistical error terms.

4.2 Selecting the data source and the type of bonds

It is important that a data source is selected that can deliver the required data fields of high-quality information about the bond market population in a reliable and timely manner covering the period under review. These data sources are, in many cases, institutional or commercial data vendors, which collect, repackage and redistribute financial market news, and reference and price data. Sample test data are necessary to test the data sources' ability to satisfy the availability of the required reference data (static) and daily price data (dynamic) of euro area government bonds, as well as being compliant with the available technical infrastructure and procedures. In the testing exercise carried out with the assistance of the ESCB, the bond data were provided by EuroMTS Ltd and covered two years of daily bond price data. Where ratings are applied, these are provided by Fitch Ratings. The raw dataset covered approximately 720 euro area government bonds and, following the application of the conceptual framework, a high-quality bond population was selected containing a daily average of approximately 320 euro area government bonds.

4.2.1 Sectors and credit risks

A good statistical starting point is to conceptualise the bond population, before extracting a sample and analysing the sample representation. A conceptually sound sector breakdown is required, preferably in line with international standards, e.g. the ESA 2010 classification[13], which distinguishes primarily between five types of issuers (ESA 2010 code numbers are given in brackets):

- Central government (excluding social security funds) (S.1311);
- "Other general government" denotes state government (S.1312), local government (S.1313) and social security funds (S.1314);
- Monetary financial institutions (including the Eurosystem (the ECB and the national central banks of the euro area) (S.121), deposit-taking corporations (S.122) and money market funds (S.123));
- Other financial intermediaries (S.125);
- Non-financial corporations (S.11).

It is indeed informative to study and compare yield curves of different sectors taking the different credit rating classes into account. For this purpose, the focus is on selecting debt securities issued by euro area central governments (S.1311). Central

[13] The European System of National and Regional Accounts (ESA 2010).

government bonds with special features, including those with specific institutional arrangements, are excluded[14] as well as variable coupon bonds, including inflation-linked bonds, perpetual bonds, convertible bonds and bonds with embedded options. State and local government bonds may be considered for national yield curves depending on availability and the size and activity of the respective local bond market, but are not included for the euro area as they may present special features tailored to local markets. According to the ECB Securities Issues Statistics, the general government sector represents 48% of all amounts outstanding of securities issues by euro area residents, with a total nominal value of EUR 7,950 billion outstanding as of October 2017. The focus on central government bonds with a high credit rating is important as this will provide the central banking community, and others, with a measure of the term structure for perceived risk-free instruments and a measure of the maturity structure of (ex post) minimum funding costs of governments.

Alternatively, one could select only securities with benchmark status. Using only benchmark bonds may lead to a more stable composition of the sample (over time)[15] and may provide the lowest level of representation of a yield curve; however, on the other hand, this may mean that only bonds at pre-fixed (preferred) maturity points rather than from the full maturity spectrum are selected and will also lead to a significant reduction in the sample population[16]. Furthermore, a precondition for selecting only benchmark bonds is to define a conceptually sound (euro area-wide) definition of benchmark bonds. Currently, to select a benchmark bond, the bond with the lowest yield within a certain maturity range or the bond which is most frequently traded is identified, or a price process is used to identify the bond which best represents the common movement of the entire market.[17]

4.2.2 Liquidity considerations

Liquidity considerations play a major role in the selection of bonds to be included in the testing sample. The more liquid a market is, and the more trades that are done, the better the information content of prices in this market segment is. A direct measure of liquidity would be to use turnover[18] values or bid-ask spreads. A more indirect measure of market liquidity might be to apply a minimum quantitative threshold of the amount outstanding of individual bonds. This indirect measure is not recommended – although it is easier to apply – as it would remove all small issuance series, which in many cases are illiquid, but does not provide sufficient evidence that the remaining bonds are liquid. Furthermore, this method may also exclude liquid bonds in small domestic financial markets.

[14] For example, Brady bonds, convertible bonds and bonds with non-regular structures (bonds with embedded option calls and puts and step-up/step-down bonds).

[15] This may be the case for national yield curves. The empirical evidence from the euro area testing exercise demonstrates that the sample population of the (highest or all-rated) euro area government bonds is stable over time.

[16] In our dataset, the sample population would be reduced by 50% if only benchmark bonds are selected.

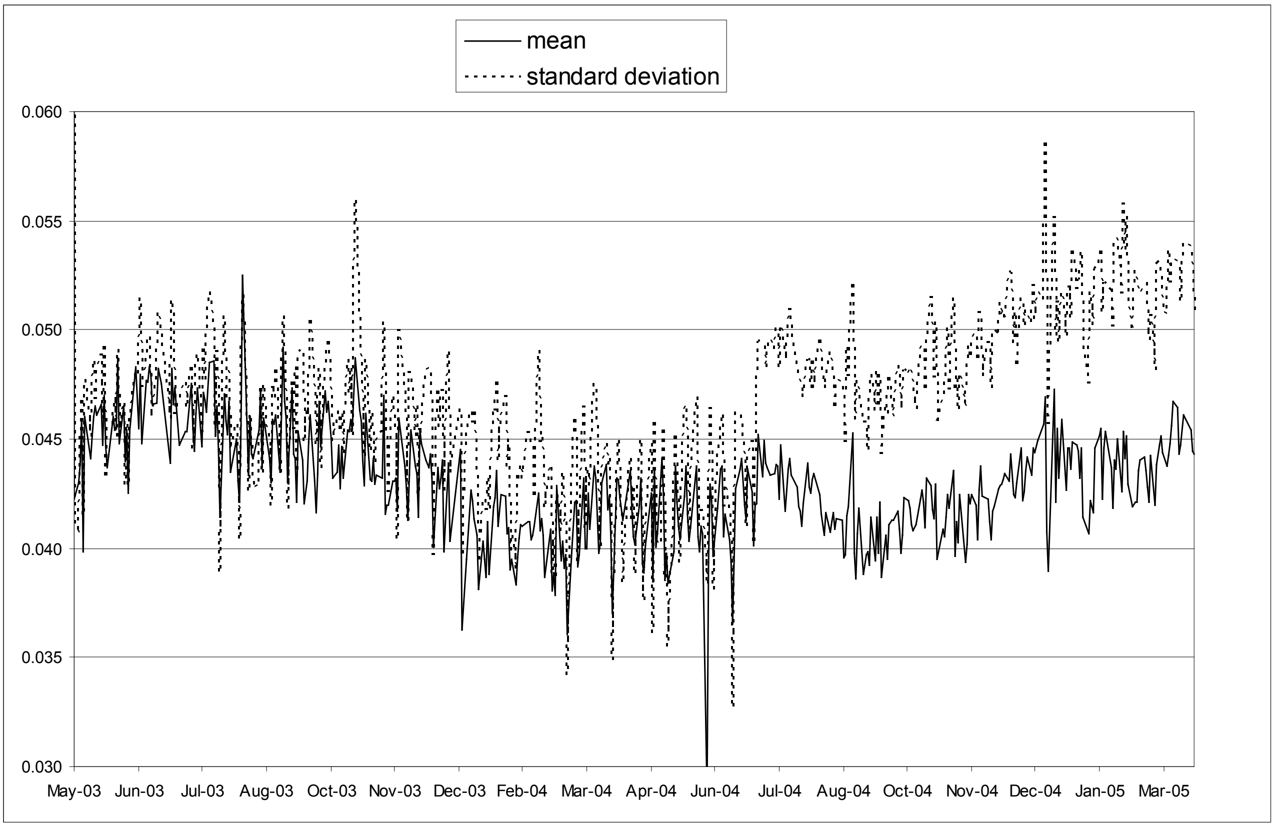
[17] See Dunne (2004).

[18] Measured as the volume of individual daily trades.

Given today's wide range of available data sources, turnover data on individual bonds are frequently available. It is considered reliable price information[19] if the minimum trading volume is above EUR 100,000-200,000 per day[20]. Within this testing exercise, the threshold level was set significantly higher (by a factor of 5-10) to EUR 1 million per day and only bonds with an average turnover of EUR 1 million were selected in the data sample (see the "Price information" section below for further considerations). A bid-ask spread threshold can also be applied as an alternative to, or a supplement to, turnover data, as shown in Chart 2.

Chart 2

Example of mean daily bid-offer spread of the bond population



Note: data: May 2003 to March 2005.

From Chart 2 above, the mean bid-offer spread is relatively stable during the two years of daily data and varies around 4.0-5.0 basis points, with a slightly increasing standard deviation during the latter part of the period.

19 This depends on the structure of the financial markets and the number of actual trades.

20 Average value over a limited period of trading days.

4.2.3 Price information

It is accepted and well-known that prices from data vendors differ for the same financial instruments, as prices are expressed for instance in terms of: (a) executable pricing (binding quote); (b) reference pricing (non-binding quote); (c) real-time pricing (reflecting the price of a single trade at that point in time); (d) mid-pricing (average of "best" bid-offer rates within a fraction of the trading day); (e) pricing according to the data vendor's underlying supplier of the price information (different trading facilities, stock exchanges or OTC contributor from single or multiple banks); and (f) the time the price was recorded (open, mid or close of market), among others[21]. Therefore, the daily time stamp plays an important role when collecting the prices of bonds, as significant intraday price volatility may occur during a trading day, and when clarifying the pricing concept of the data source. The statistics presented in the following tables and graphs are used to select additional bonds for the sample (or to remove bonds from the sample population).

Chart 3

Percentage of trades and quotes of government bonds during the two years of daily price data

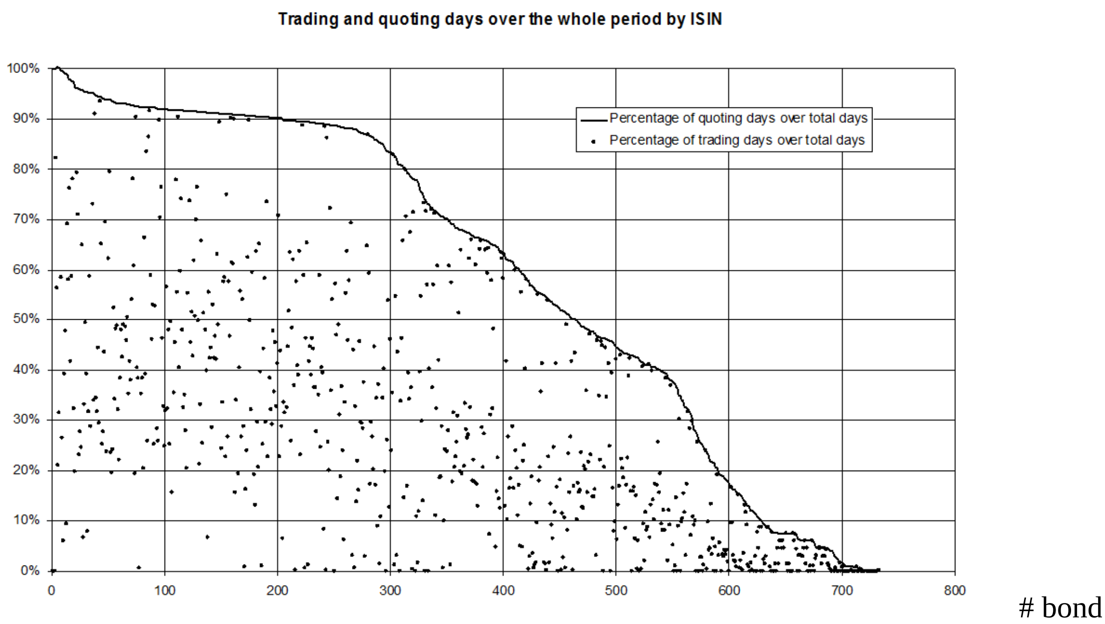


Chart 3 shows two types of information. The first line (thick line) represents the percentage of bonds that has been quoted during the 502 actual trading days (two-year data period). For instance, it can be seen from the graph that approximately 200 bonds (of the total sample of 720 bonds) have been quoted on more than 90% of the trading days.

[21] Commercial data vendors have expanded lists of different price options.

The second type of information is shown by the dots in the graph. Each dot represents an individual bond and describes the percentage of trading days on which the bond has actually been traded over the 502 trading days, even though the bonds (executable) are quoted during non-traded days.

Against this background, it should be noted that not all bonds in the population are actually traded every single trading day and therefore consideration should be given to defining an acceptable threshold for including bonds, depending on the frequency or number of sequential trading days. Note that the statistics shown in the graph above include bonds which have matured during the two-year period.[22]

To define a threshold level, consideration should also be given to the type of price information provided by the commercial data vendor as described above. In this data sample, the bid-offer prices are executable pricing; therefore, the difference between a real trade price and an executable price is minimal, if any. The reason for this is that executable prices[23] are binding for the offer and are immediately executable in the trading system. In fact, the difference between the real traded prices and quotes close to 17:00 can vary daily – on average between 1% and 1.8% (see Table 1). Furthermore, the number of average daily traded bonds with real traded prices close to 17:00 varies depending on the domestic market from [0.1 to 25.9] bonds in March 2005, whereas the number of quoted bonds varies within the interval of [5.0 to 81.7] bonds. This implies that if real traded prices are selected, the sample size will be significantly reduced (see Table 2). On the other hand, if mid-price quotations are selected, there is a clear reference time near the market close of 17:00 attached, which provides a significantly larger sample population with quotes very similar to actual prices and, importantly, all quotes have the same reference time stamp.

If, however, there is a large difference between the traded and quoted pricing, it may be useful to define such a quantitative threshold for the inclusion/exclusion of bonds, for instance, that each bond should be traded at least three out of five sequential trading days in order to be included in the bond sample.

22 With reference to Table 4, one can calculate that, on average, 26% of the sample population has a residual maturity of less than or equal to one year.

23 Executable prices are in this case based on the “mid-price”, which is defined as a flat price quote based on the average of the best bid-ask prices at or before 17:00 with a maximum spread of three basis points.

Table 1
Differences between traded prices and quotes in March 2005

(differences as a percentage of price; averages of daily data)

	Traded bonds				
	total	after 15:00	before 15:00	after 16:00	before 16:00
AUSTRIAN	-0.05	-0.04	-0.07	-0.01	-0.06
BELGIAN	-0.04	-0.03	-0.04	-0.01	-0.04
DUTCH	-0.04	-0.02	-0.05	-0.02	-0.04
FINNISH	-0.03	-0.02	-0.03	-0.02	-0.03
FRENCH	-0.03	-0.01	-0.04	-0.01	-0.04
GERMAN	-0.04	-0.02	-0.05	-0.01	-0.05
GREEK	-0.03	-0.02	-0.04	-0.01	-0.04
IRISH	-0.06	-0.01	-0.07	-0.01	-0.07
ITALIAN	-0.02	-0.01	-0.04	0.00	-0.03
PORTUGUESE	-0.03	-0.02	-0.05	-0.01	-0.04
SPANISH	-0.04	-0.03	-0.04	-0.01	-0.05
Total	-0.04	-0.02	-0.05	-0.01	-0.04

Table 2
Number of bonds quoted/priced during March 2005

(averages of daily data)

	Market code	Quoted bonds	Traded bonds	Traded bonds (after 15:00)	Traded bonds (after 16:00)
AUSTRIAN	ATS	12.6	6.5	1.9	0.8
BELGIAN	BEL	39.8	16.6	6.0	2.8
DUTCH	NLD	20.7	9.1	3.0	1.5
FINNISH	FIN	6.0	5.0	1.6	0.6
FRENCH	FRF	81.7	24.6	7.8	3.4
GERMAN	GEM	48.1	18.9	6.3	3.4
GREEK	GGB	19.0	14.7	6.1	3.0
IRISH	IRL	5.0	1.7	0.3	0.1
ITALIAN	MTS	57.9	53.8	39.8	25.9
PORTUGUESE	PTE	16.0	10.3	5.2	2.6
SPANISH	ESP	42.0	10.7	4.2	2.0

In line with the objective of gauging market expectations from yield curves, the underlying yields/price should reflect a homogeneous information set. This requires yield or price data to be taken at about the same time. The dataset used for the estimation of yield curves comprises the close-of-market prices, i.e. the prices of the last executed transactions at 17:00 Central European Time (CET) for each bond. If there have been no transactions, quotes posted at that time are used.

4.2.4 Constructing the maturity spectrum

The selection of the maturity spectrum to represent the yield curve needs consideration. One can construct a yield curve covering the full maturity spectrum or construct two or more yield curves according to the type and characteristics of the available underlying instruments. In principle, one curve could be constructed by using financial instruments with a residual maturity below one year and another curve covering government bonds with residual maturities above one year[24]. For maturities below 12 months, money market rates can be considered (overnight to 12-month maturity), with liquidity considerations taken into account. In this case, this would be EONIA rates (the effective overnight reference rate for the euro) and EURIBOR rates (the money market reference rate for the euro).[25] The spread between EURIBOR interest rates vis-à-vis government bond yields with similar residual maturities is shown in Table 3.

Table 3

Differences between average bond yields and EURIBOR rates at different maturities[26]

(sample period: 1 April 2003 to 31 March 2005)

	Mean	Min	Max	St. Dev.	Median	4th-1st quartile	1st percentile	99th percentile
1 month	-0.039	-0.098	0.074	0.013	-0.037	0.011	-0.054	-0.008
2 months	-0.046	-0.086	-0.024	0.009	-0.045	0.011	-0.058	-0.030
3 months	-0.049	-0.076	-0.022	0.007	-0.049	0.008	-0.057	-0.030
4 months	-0.051	-0.077	-0.012	0.007	-0.051	0.008	-0.058	-0.031
5 months	-0.051	-0.074	-0.011	0.006	-0.050	0.008	-0.058	-0.037
6 months	-0.051	-0.094	-0.010	0.007	-0.051	0.008	-0.059	-0.034
1 year	-0.044	-0.088	-0.007	0.011	-0.044	0.013	-0.059	-0.016

The daily differences between the average bond yields and EURIBOR rates at various maturities show that yields of bonds with residual maturities of one and two months tend to display slightly abnormal behaviour in terms of standard deviation, inter-quartile difference and 99th percentile of the distribution versus EURIBOR rates, which is marked in grey in Table 3 above.

The population of debt securities with a residual maturity of below one year represents 20-30% of the total number of debt securities (see Table 4).

24 This split is set at one year, which reflects the borderline between the money market segment and the capital markets segment.

25 www.emmi-benchmarks.eu

26 For each date and class of residual maturity, the average yield for all bonds belonging to the corresponding class of residual maturity is first calculated; then, the difference between this average yield and the corresponding EURIBOR rate is taken. The statistics shown in the table relate to the distributions (one for each class of residual maturity) of these differences throughout the two years of sample data.

Table 4

Three snapshots of the share of short-term securities by maturity bracket

(percentage composition)

	0-3 months	3-6 months	6-9 months	10-12 months	more than 1 year	Total
15-Apr-03	8	10	6	7	70	100
15-Apr-04	5	5	6	4	79	100
15-Mar-05	7	9	5	6	73	100
Average of daily data from 01/04/03 to 31/03/05	7	8	6	6	74	100

Charts 4 and 5 present the volatility of yields across bonds within different residual maturity brackets, where the volatility of the lowest residual maturity brackets is significantly higher than the volatility of any other classes of maturity. If the lowest maturity bracket is removed from the sample, the volatility of the three residual maturity brackets (3-6 months, 6-9 months and 9-12 months) has a fairly stable and normal behaviour (see Chart 5).

Chart 4

Volatility of yields across bonds by class of residual maturity

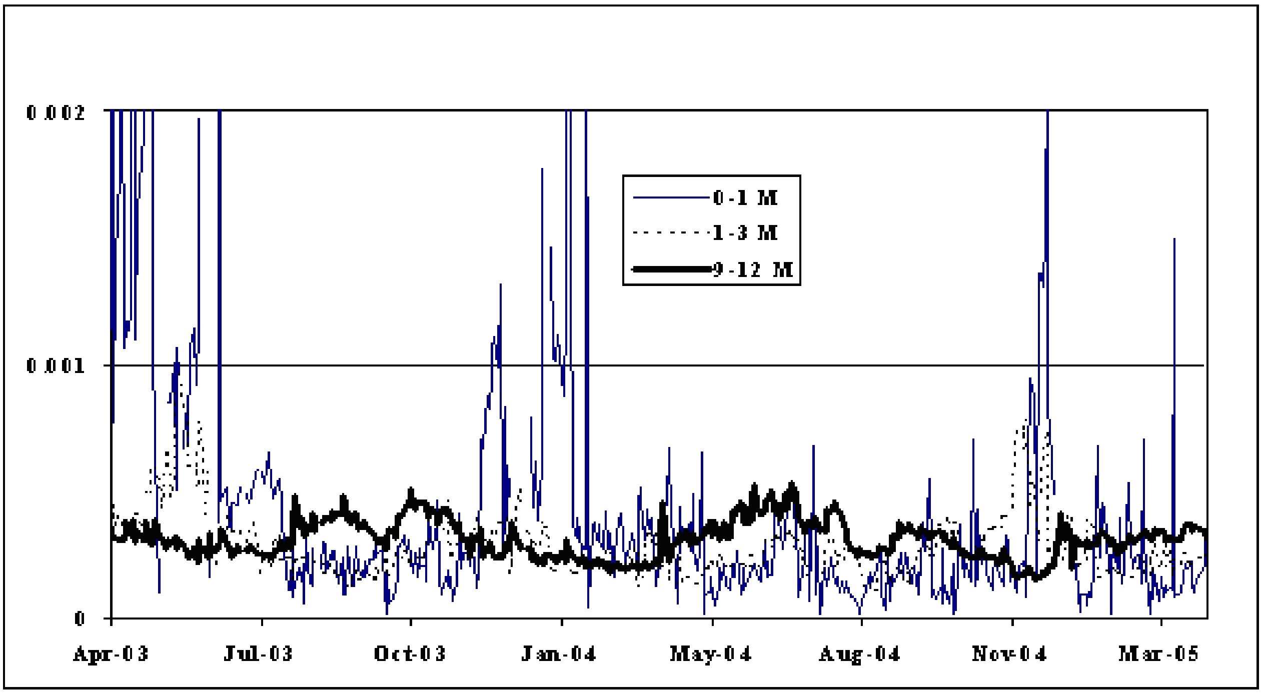
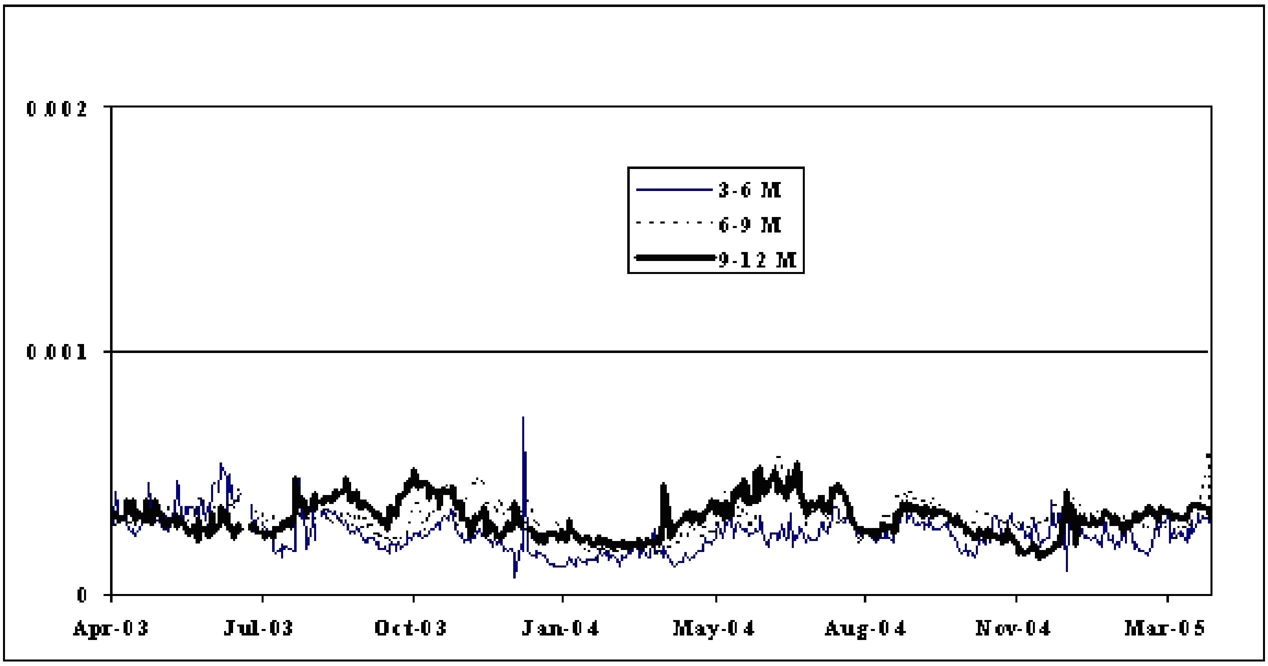


Chart 5

Volatility of yields across bonds by class of residual maturity



Following the descriptive statistics above, bonds with less than three-month residual maturity should be removed from the sample population as they exhibit significantly higher volatility than other bond maturity classes. These bonds demonstrate abnormal behaviour vis-à-vis market interest rates and would then be a potential cause of bias in yield curve estimations, if maintained in the sample population.

On the other side of the spectrum, only a few bonds have a remaining maturity exceeding 30 years and often their price mainly reflects the exceptional demand of institutional investors that need assets with a long duration. In view of these possible distortions and their lack of liquidity, these very long-term bonds should also be removed.

Therefore, following the statistical tests on the bond maturity profile, it seems prudent to construct one government yield curve using debt securities with residual maturities above three months and below 30 years. Furthermore, it seems advisable not to supplement the bond population with money market instruments, as there is a sufficient amount of bonds represented in the 3-12 month maturity spectrum. The advantage of this is that the resulting curve will represent a pure euro area government bond curve, which will enhance the explanatory power of the informational content of the curve, without introducing undesirable biases in the model estimations by mixing capital market instruments with money market instruments.

4.2.5 Removal of outliers and stability of the sample population over time

Despite the application of the various selection criteria, the yields of a few bonds may still deviate significantly from the rest. To prevent noise in the yield curve estimation, these outliers must be removed from the sample. Outliers are traced separately for a number of residual maturity brackets. Bonds with yields that deviate by more than two standard deviations from the average in each maturity bracket are considered to be outliers and can be removed from the sample. In each of these brackets, the average yield and standard deviation are calculated. This procedure is iterated in order to reduce the sensitivity of the analysis to potentially large outliers eliminated in the first round that could have distorted the average yield level and the standard deviation (see Chart 6). Applying the above procedure to the two years of daily data provides a master population of between 269 and 354 daily qualifying bonds with a daily average of 318 bonds (see Chart 7).

Chart 6

Example of outlier detection

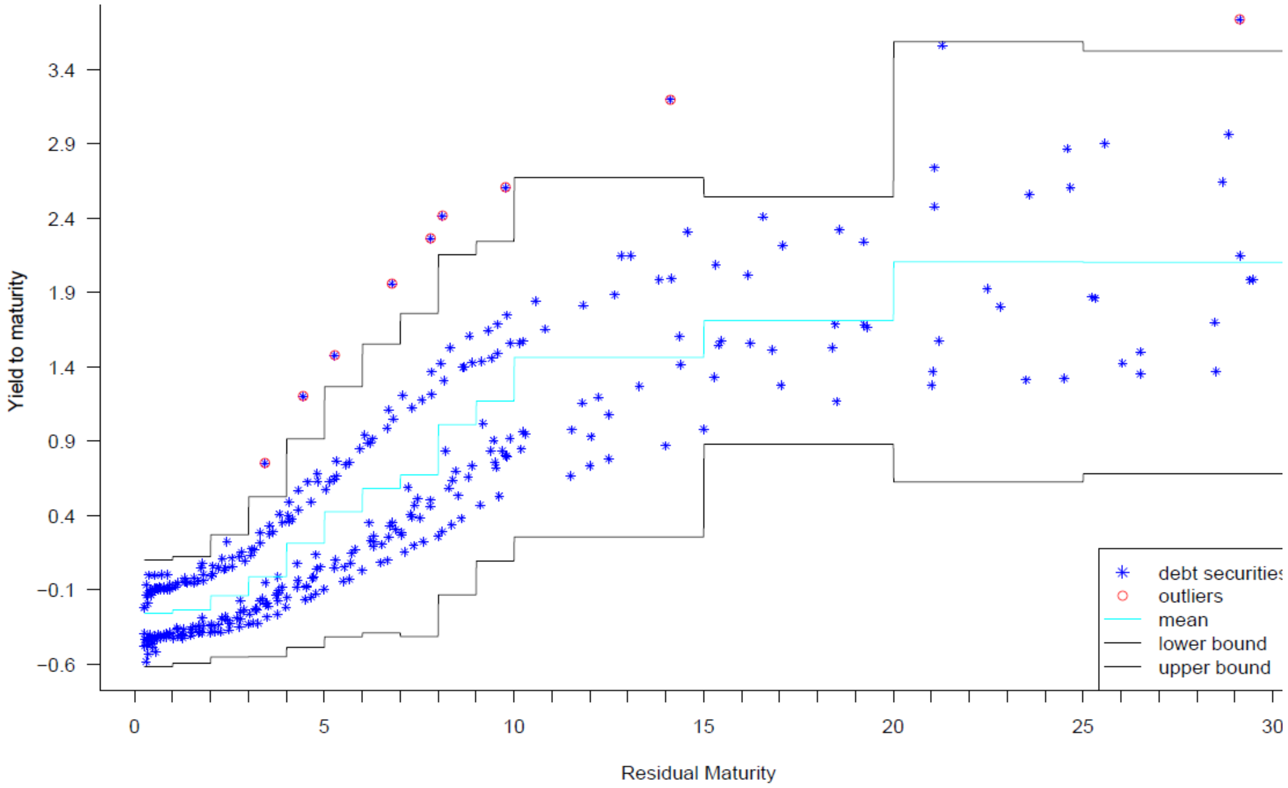
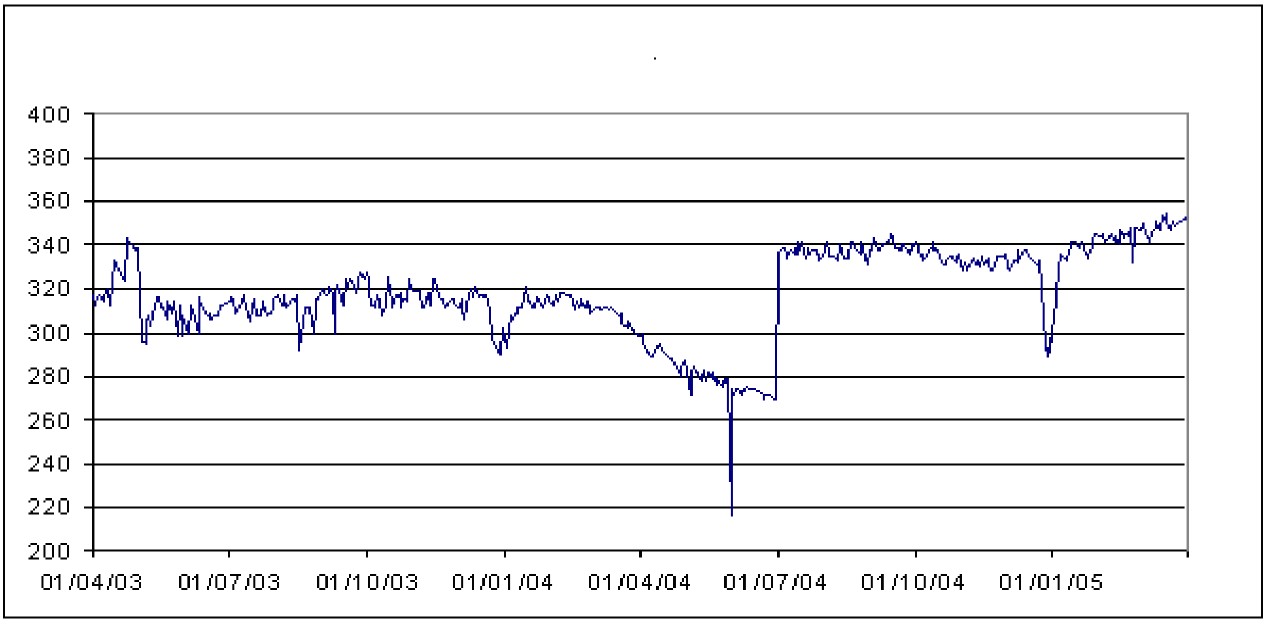


Chart 7

Total number of selected bonds per day after applying the conceptual framework



Furthermore, the composition of the basket of euro area bonds is considered to be large enough so as not to be seriously impacted by individual national public holidays. The increase in bonds (+65) between 30 June 2004 and 1 July 2004 is mainly due to the inclusion of zero-coupon bonds from Belgium (+17), France (+13) and Italy (+14). Chart 6 also demonstrates that the sample population increases over the two-year period, which could create a comparability issue regarding the stability and representativeness of the sample population over time, in particular if certain residual maturity spectrums are neglected or not represented. Therefore, Charts 8 to 10 show the distribution of bonds over ten maturity classes, where (on average) the lowest numbers of bonds can be found for the maturity spectrums of between five and ten years. These maturity brackets have an average of approximately 16 bonds and are quite stable throughout the two-year period. The maturity bracket with the highest number of bonds can be found at the very short end of the market, where approximately 66 bonds (on average) are shown in the three months to one year residual maturity bracket.

Chart 8

Number of bonds per maturity category below four years

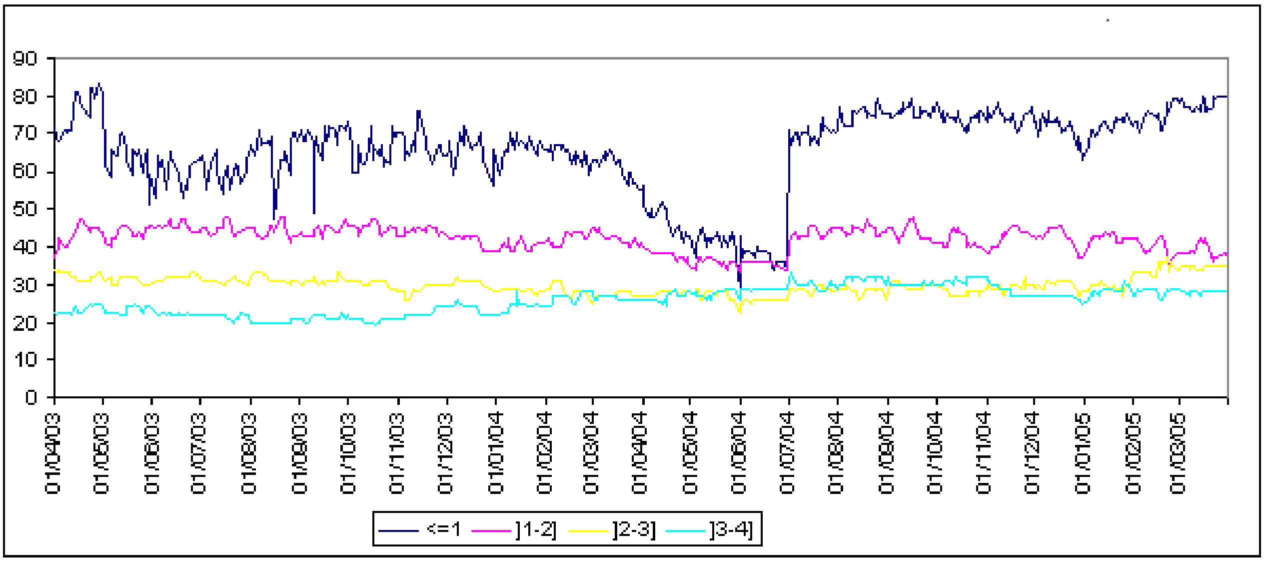


Chart 9

Number of bonds per maturity category of four to eight years

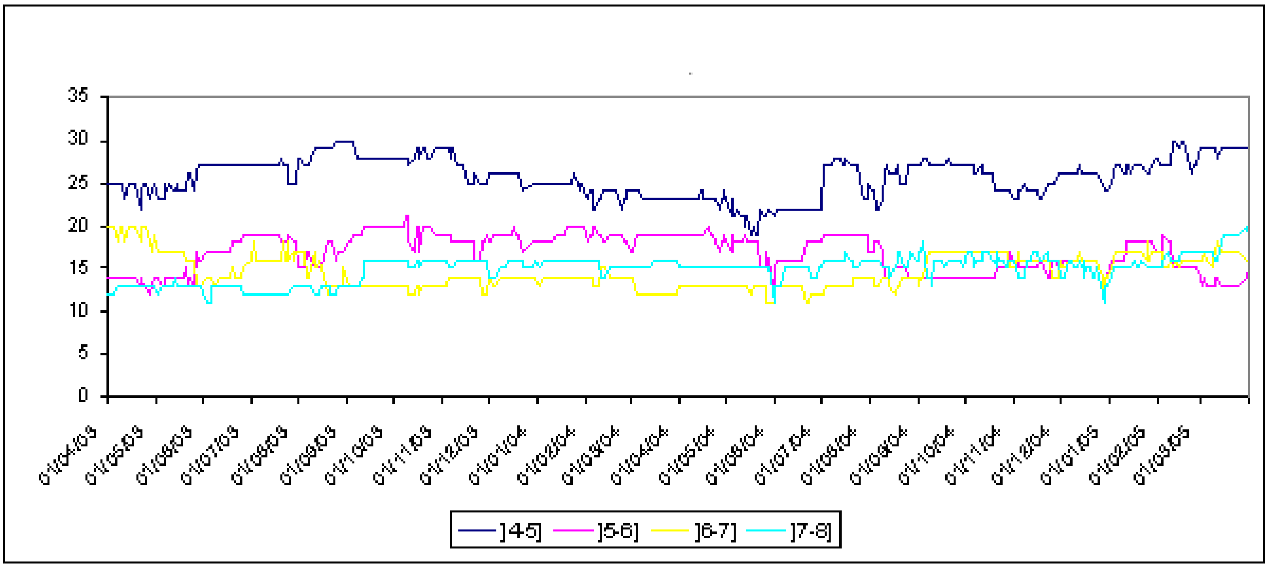
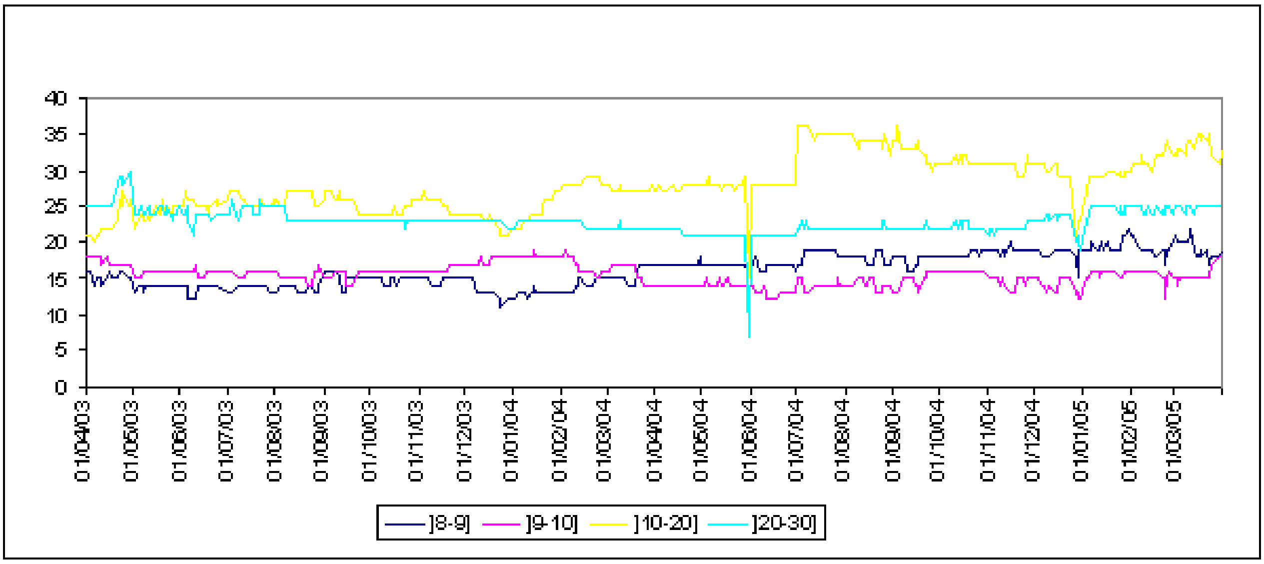


Chart 10

Number of bonds per maturity category of eight to 30 years



Furthermore, there seems to be a reduction in the sample size on 24, 29 and 31 December, which reflects the Christmas season.

Following the descriptive statistics, it can be concluded that the sample size is sufficiently large, also considering the various national holidays, and that there is a sufficient number of bonds represented in all maturity spectrums. A yield curve can therefore be produced covering all business days following the target calendar[27].

4.2.6 Adjustments for coupon effects and tax

There is a positive relationship between coupon and price and an inverse relationship between price and yield. The extent to which changes in bond prices affect the yield is weighted by the coupon value. If the coupon is significantly different across countries, both the yield and its changes will not be comparable. In practice, there is no suitable way of directly adjusting for coupon effects. Similarly, due to its complexity, no tax adjustments are considered. Therefore, in the interests of comparability, yields before tax are used. However, the population of bonds is traded in different markets with different trading calendars and market conventions. Adjustments are therefore made to guarantee comparability, for example regarding settlement and day-count conventions.

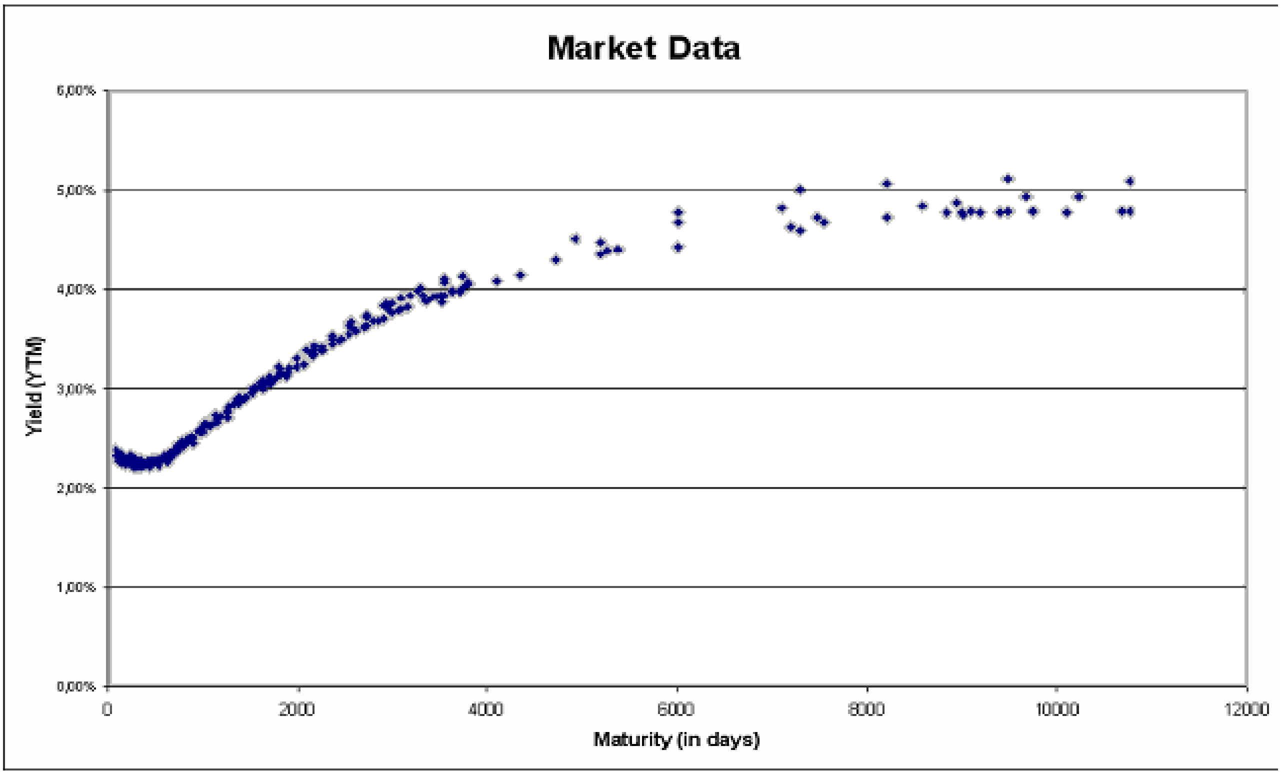
[27] This excludes the three falls on 24, 29 and 31 December.

4.3 Presenting the final high-quality bond sample

Applying this selection process produces an average sample population of 318 euro area government bonds. Chart 11 below plots the 284 observations for 9 May 2003. As expected, it shows a high concentration of bonds in shorter maturities and a much lower representation of bonds in the longer-term maturities. This observed phenomenon may have an impact on the model estimations and subsequently on the error statistics, as relatively more bonds in shorter maturities will bias the error statistics since the different models more precisely fit the intervals in which the information is better. This can be compensated for by using a weighted indicator for the goodness-of-fit statistics, such as the weighted mean absolute error (WMAE)[28].

A visual inspection of the plot of the selected sample bonds (see Chart 11) already indicates a relatively smooth yield curve. Charts 12 and 13 show the distribution of bonds per country and per maturity spectrum before and after applying the conceptual framework to the bond sample.

Chart 11
Plot of the high-quality dataset for 9 May 2003



[28] The weighted mean absolute error equals the average distance between the actual yield and the estimated curve, using the inverse of the square root of duration as the weighting factor. See the subsequent sub-section on "Evaluation criteria for the yield curve testing exercise" for further details.

Chart 12

Distribution of bonds per country before and after applying the conceptual framework

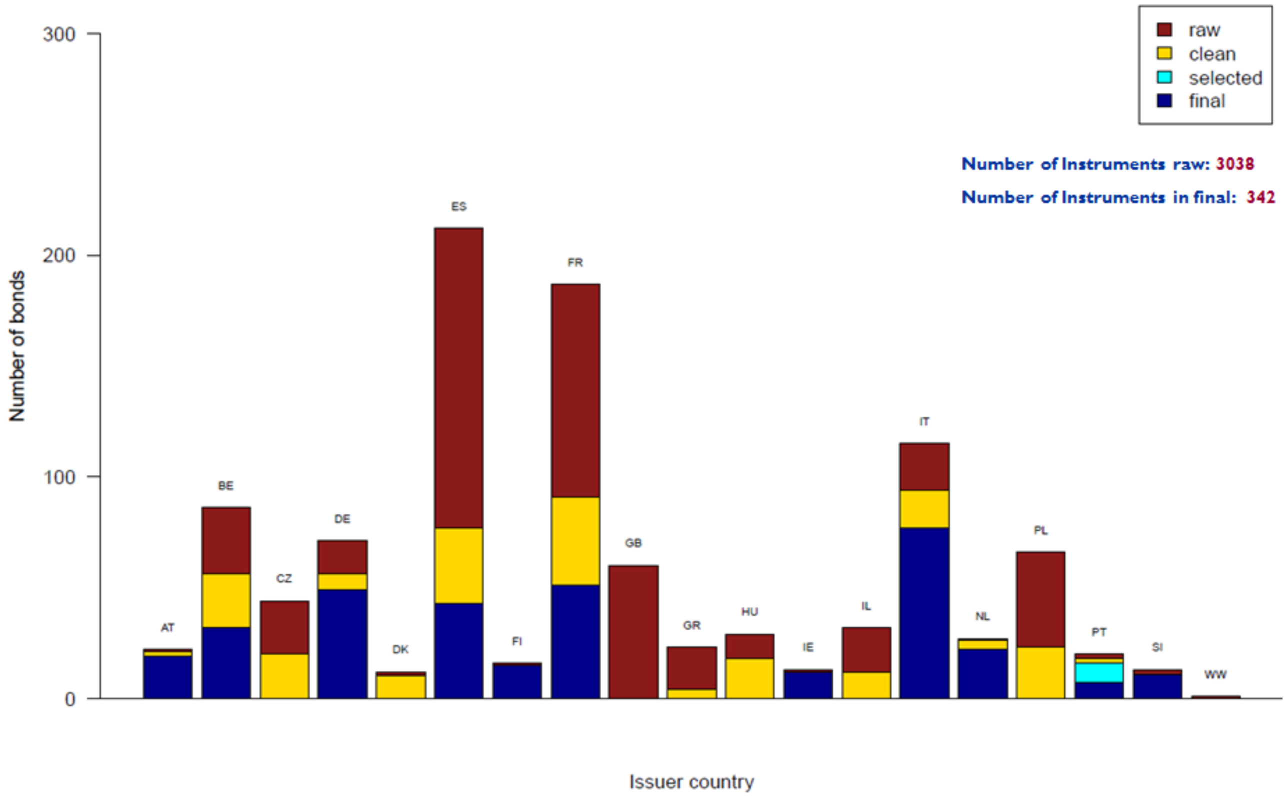
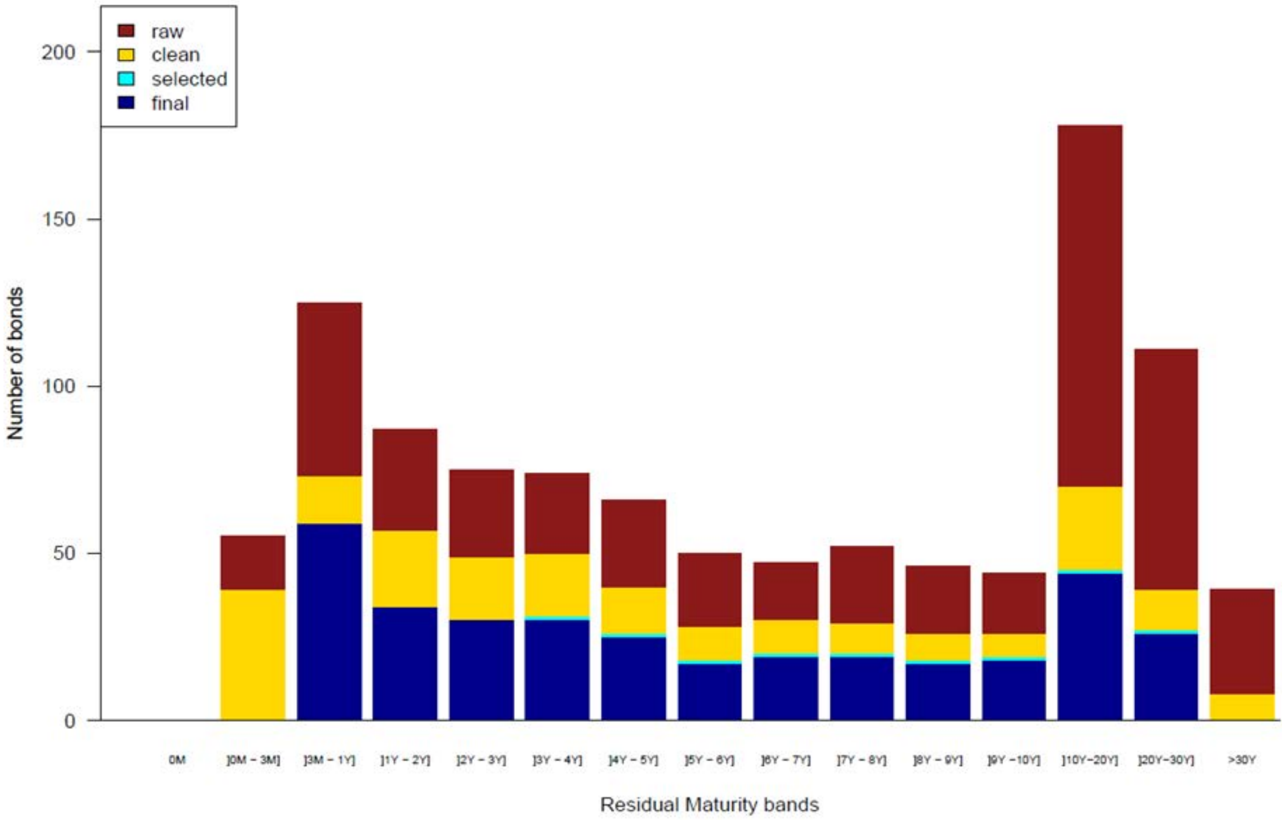


Chart 13

Plot of the distribution per maturity bracket before and after applying the conceptual framework



A high-quality sample of debt securities which is relatively stable over time has now been obtained. This quality sample can now be used for model testing. Before doing so, the next chapter will prepare the quality sample for model testing and specify the statistical tests that can be applied as part of the assessment of the performance of the selected yield curve models.

5 Preparing the empirical exercise

Following the creation of a high-quality dataset in Chapter 4, Chapter 5 will focus on dividing the dataset into several subsets as part of preparing the yield curve testing exercise. A dataset for the in-sample testing and another for the out-of-sample testing are selected and sub-datasets reflecting periods of different market expectations and conditions are created. In this chapter, the statistical tests to be used in the subsequent chapter to evaluate the performance of the selected yield curve models are also selected. As in the previous chapters, a few illustrative examples are included to facilitate the preparation process.

5.1 Selecting an in-sample and out-of-sample population

Bliss (1996) demonstrates that using the in-sample goodness of fit as the sole criterion for judging term structure estimation methods can be misleading.[29] The non-parsimonious methods such as the simple bootstrapping (considered in the Bliss paper) and the spline-based method can fit the observed data with arbitrary precision. Thus, by relying only on the in-sample fitting statistics, these models tend to always perform better than the Svensson or Nelson & Siegel model. However, once the spline-based models are tested with out-of-sample securities, they may perform quite poorly, or react sensitively to changes in the estimation sample. Therefore, both in-sample and out-of-sample statistics should be used as two separate evaluation criteria for the testing exercise. In principle, if the emphasis were placed on flexibility, one would favour and attach more weight to the results of the in-sample statistics, whereas if robustness were to be emphasised, more weight would be given to the out-of-sample statistics.

The out-of-sample selection should include bonds covering the full residual maturity spectrum from three months to 30 years of residual maturity and be sufficiently small so as not to influence the fitting of the curves. Therefore, all bonds in the selected sample are ranked by ascending order of residual maturity, and one out of ten bonds is selected for the out-of-sample population. In this case, we have an in-sample dataset consisting of an average of 286 bonds and an out-of-sample bond population of an average of 32 euro area government bonds.

5.2 Comparison of different time periods for model testing

It is important to test the performance of the different models during different market conditions and expectations, as models may not necessarily perform symmetrically during periods of market expectations of interest rate hikes or cuts, or periods of high

29 Bliss, R.R. (1996), “Testing Term Structure Estimation Methods”, Working Paper 96-12a, Federal Reserve Bank of Atlanta.

volatility on the bond markets with increasing credit spreads. Against this background, five periods are selected and tested separately.

(a) Summary statistics covering the full time spectrum

(b) Summary statistics covering the time period from 1 June 2003 to 7 June 2003

During this period, there were strong interest rate cut expectations. The ECB Governing Council decided to lower official interest rates from 2.50% to 2.00% with effect from 6 June 2003. The period represents expectations on the day before the Governing Council meeting and the day after the decision of the Governing Council took effect.

(c) Summary statistics covering the period from 15 November 2004 to 31 January 2005

During this period, the euro reached a high level, after a very quick rally (13% in a month against the US dollar), and the ECB/Eurogroup/European Commission reacted with a joint speech (6 December 2004) to prevent speculation. Secondly, the strong economic outlook and the rise in oil prices helped the ECB to adopt a hawkish view (7 December 2004), even warning about credit tightening: the market anticipated a rise of 25 bps, and yield levels were rising. Finally, the rates were left unchanged, since the rise in the euro was almost equivalent to a 30-50 basis point increase in interest rates (20 December 2004, O. Issing): market expectations turn out to be wrong.

This period offers the possibility of analysing the trend in short-term interest rate expectations, monetary policy expectations, inflation expectations and exchange rate expectations. In addition, the curve is flattening during this period.

(d) Summary statistics covering the period from 1 February 2005 to 31 March 2005

During this period, a widening of credit spreads (as several large corporate credit ratings were downgraded, such as General Motors and Ford) was observed, at the same time as the reform of the Stability and Growth Pact and the referendum on the EU constitution in France.

(e) Summary statistics covering the individual days of 12 August 2003, 9 December 2003, 22 June 2004, 21 September 2004 and 15 March 2005

These individual days are some of the days in which the slope of the yield curve experienced the biggest changes compared with the day before.

5.3 Evaluation criteria for the yield curve testing exercise

It is important that all yield curve models undergo exactly the same statistical tests and are compared according to their performance. The evaluation criteria provide an objective measure of how well each of the models captures the shape of the underlying term structure and – at the same time – fits the dataset required for the smoothness and flexibility test. The following three evaluation criteria are proposed.

Flexibility & goodness-of-fit criterion – this criterion aims at capturing movements in the underlying term structure, in particular for shorter maturities. For this purpose, a goodness-of-fit test can be applied. A test is performed to see whether the estimated curves can accurately price a bond which has not been used to estimate the curve. For this purpose, the "out-of-sample" goodness-of-fit test is used and the means and standard deviations are compared with the in-sample results (see box below). In addition, two measures of pricing errors can be applied, namely the weighted mean absolute error[30] and the "hit rate"[31]. See Appendix 1 for the formulae of error statistics.

Robustness criteria – this criterion measures the sensitivity to price changes and provides a way of ensuring that changes in the data at one maturity (such as at the long end) do not have a disproportionate effect on other maturities (such as at the short end). The principle is to change the underlying bond prices marginally and compare the original curve with the marginal curve and measure how robust the estimated curve is to marginal price changes.

Smoothness criteria – this criterion aims to provide a recommendation on which models provide relatively smooth curves for the purpose of supplying a market expectation curve for monetary policy purposes. The trade-off and slight preference is to ensure better data without trying to fit every data point. The degree of smoothness is tested based on spreads between different maturities to assess if the spread variation indicates changes in the slope, which allows us to determine the behaviour of the curve over time.

Box
Explainer for the statistical tests

These statistical tests are often referred to as diagnostic checks; for instance, one can use the "goodness-of-fit" test to diagnose how good a statistical model is at predicting an observed value. Or, in other words, do the observed values tend to lie close to, or far away from, the curve. The difference between the predicted value and the observed value is labelled the "residual" and if these residuals behave in a consistent manner (for instance, within a certain range), one can say that the model is appropriate to be used for estimation. To compensate for the fact that there are

30 This is a measure of the average distance between the actual yield and the estimated curve, using the inverse of the square root of duration as the weighting factor. As a result, yield errors must be scaled to deflate the influence of the less precise observation (in this case where the data gaps are wide, such as between ten and thirty years).

31 The hit rate represents the percentage of observed yields which lie within a predefined spread from the curve, thus indicating that models with a higher percentage level of hit rates fit the dataset better than models with a relatively lower percentage level.

significantly less bonds within the longer maturity bands (for instance, within the maturity band of 10-30 years), a weighting factor can be applied. In the tests, the inverse of the square root of "duration" is used as the weighting factor. Similarly, if the weighted residuals behave in a consistent manner, it is possible to say that the model is appropriate. Chapter 5.3 entitled "Introduction to time series and forecasting" of Brockwell and Davis (2016) provides further examples.

The "hit rate" aims to represent a measure of how well the model fits the observed yields throughout the maturity bands. It draws up a fixed bandwidth surrounding the estimated curve and calculates the percentage of observed yields which lie within the bandwidth for each maturity band. 1 minus the hit rate thus denotes the percentage of observed yields which is outside the bandwidth. A high percentage for the hit rate indicates a better fit and a low percentage indicates a low fit. In this notation, each observed yield becomes a binary variable; the observed yield is either inside or outside the bandwidth (see also Appendix 1). Note that the hit-rate measure does not indicate how far the observed yield is from the model estimate. This is important for the spline-based models (VRP), in particular for maturities above ten years (see also the video in Appendix 5).

Smoothness tests are mainly designed in econometrics to detect abrupt breaks in time series. These abrupt breaks can be observed by studying the pattern of the spreads between different maturities of the yield curve over time. If no abrupt breaks can be found when plotting the time series, the model is said to be smooth. In other words, the variations in the spreads over time would then indicate changes in the slope of the yield curve.

The terms robust and robustness refer to the strength of a statistical model according to the specific conditions of the statistical analysis. Robustness makes it possible to explore and test the stability of the estimates in response to plausible variations in the underlying data and/or model specifications. In other words, a statistical test is described as being robust if it is not especially sensitive to small changes in the dataset and it is largely unaffected by outliers or small departures from model assumptions (see Neumayer and Plümper, 2017).

5.4 Model selection

Different models can be applied for yield curve estimations, as presented in Chapter 3, and the trade-off between smoothness and goodness of fit must be considered as part of the selection process. It therefore seems reasonable to select both parsimonious and spline-based models and test which of the models best reflects a smooth curve but is flexible enough to capture movements in the underlying euro area term structure for monetary policy purposes.

With this in mind, two parsimonious and two spline-based models were tested with the assistance of experts from the Bank of Greece, the Bank of England, the Banca d'Italia, the Banque de France and the Deutsche Bundesbank.[32] Where the same

[32] As part of the work conducted by the ESCB Task Force on Financial Market Statistics. In particular, Vasileios Georgakopoulos and Polychronis Manousopoulos (Bank of Greece), Giovanni Guazzarotti (Banca d'Italia), Fahd Rachidy and Maryam Housni Fellah (Banque de France), Matthew Hurd, Andrew Meldrum and James Mason (Bank of England) and Jörg Meier and Jelena Stapf (Deutsche Bundesbank).

model was tested by more than one national central bank, the characteristics of the model properties differ from each other. The differences relate to the type of curve, the objective function, constraints, starting values or the algorithm applied. For instance, the Banque de France (BdF) assisted in testing a hybrid model of the Nelson & Siegel and Svensson models, which switches between the two models depending on the daily analysis of a confidence interval for the beta3 parameter value. The Svensson model is used when the beta3 value is within the confidence interval, or if not, the Nelson & Siegel model is used. This makes the BdF approach unique, and the results of the hybrid model can be compared, to some extent, with the results of the other models.

For the spline-based models, Fisher, Nychka and Zervos (1995) and Anderson and Sleath (2001) recommend that knot points are placed at approximately every three to four bonds. This method may be valid for producing national yield curves where small sample sizes are available, but will increase the risk of obtaining clustered knot points for the euro area considering the significantly larger size of available euro area bonds with similar characteristics and residual maturities. One alternative would be for the knot points to be spaced so that roughly an equal number of bonds will mature between adjacent knots. For example, the Bank of Greece applied the latter method which resulted in a knot point at approximately every sixth bond. To optimise the testing exercise, the Waggoner model is estimated using "yield minimisation", while the VRP model uses "price weighted by duration" minimisation. See Table 5 for an overview of the models tested and Appendix 3 for the model properties applied by the national central banks.

Table 5

Overview of the four selected yield curve models

Type of model	Model	Functional form[1]	Central banks
Parsimonious models	Nelson and Siegel	Exponential approximation of the discount rate function directly to bond prices	Bundesbank, Banque de France & Bank of Greece
	Svensson	Extension of the Nelson and Siegel model	Bundesbank, Banque de France & Bank of Greece
Spline-based models	Waggoner step-wise penalty function	A three-tiered step-wise linear penalty function. One step set at one year and the second step at ten years to the residual maturity level	Bank of Greece & Bank of England
	Waggoner with a variable roughness penalty (VRP) function	Continuous penalty function	Bank of England

Notes: see also Chapter 3.
1) See Appendix 2 for the detailed functional forms of the yield curves.

6 Empirical evidence and comparison of results

In this chapter, the empirical evidence and the performance of the models are reviewed by firstly looking at the results of the goodness-of-fit tests using the in-sample and the out-of-sample checks. The results of the robustness and smoothness tests are then reviewed.

A few illustrative extractions are used as examples within this chapter. The descriptive statistics, models and test results are presented in Appendix 4.

6.1 Flexibility & goodness-of-fit criterion (in-sample results)

In this sub-section, examples of the statistical results are provided, by comparing the goodness-of-fit statistics and hit-rate statistics for all periods firstly for the in-sample statistics.

As a preliminary inspection, the following yield curves may be represented by the four models in Chart 14.

Chart 14

Representation of yield curves using the four different models

(percentage points)

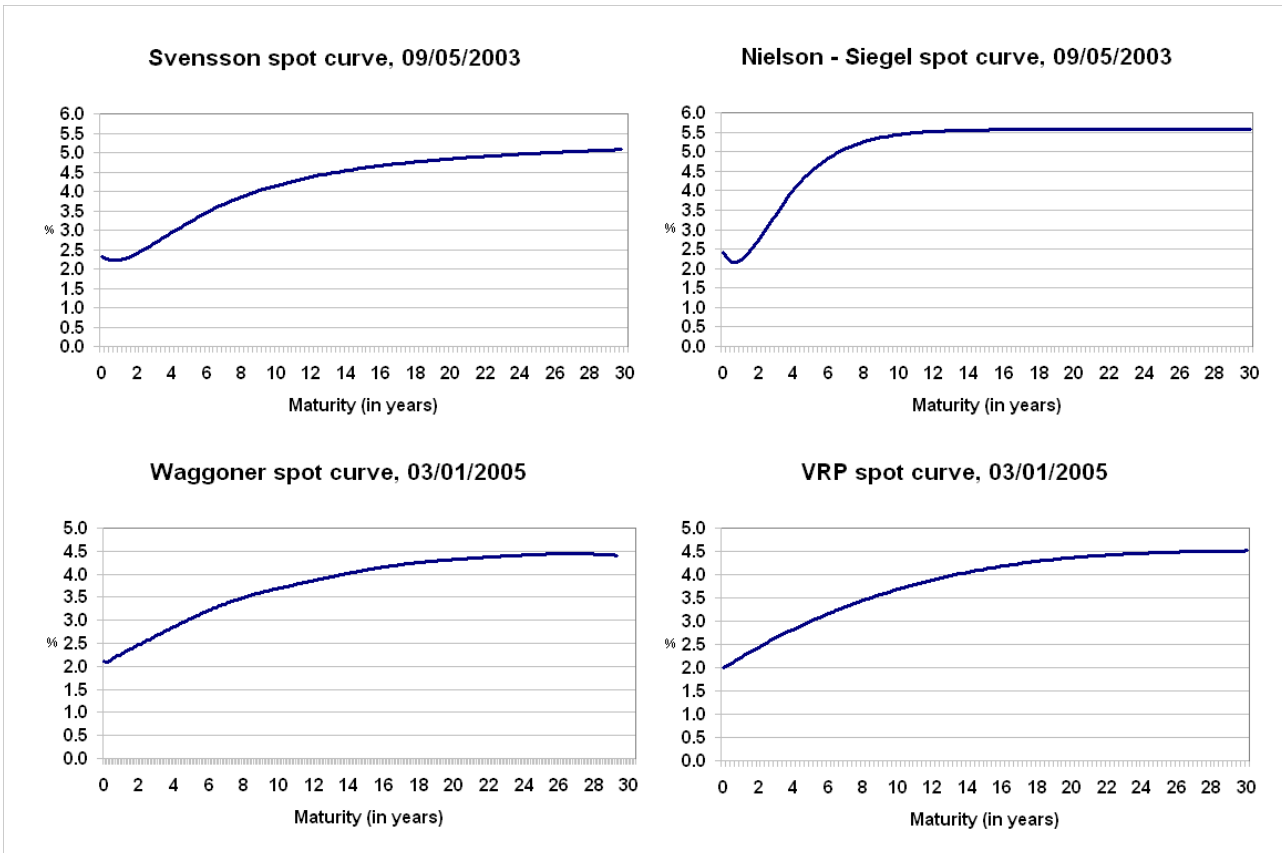


Chart 14 shows, as an example of the model, the results of the spot yield curves for the Nelson & Siegel and Svensson models (the two parsimonious models) on 8 May 2003. From a visual inspection of the shape of the two curves, it can be seen that they provide a very similar representation of the term structure of interest rates across the full maturity spectrum. Both models demonstrate a similar yield level for the short-term interest rate, the long-term interest rate and the slope and hump of the curve. Similar results are observed for the two spline-based models (the Waggoner and VRP models), where their representation is from 3 January 2005. In fact, all four models provide a very similar representation of the term structure of interest rates across the full maturity spectrum.

For the two parsimonious models, the degree of flexibility is determined by the number of parameters to be estimated, where four and six parameters are to be estimated for the Nelson & Siegel and Svensson models, respectively.

Table 6

Example of estimated parameter values for Nelson & Siegel and Svensson models

(9 May 2005)

Parameter	β0	β1	β2	β3	τ1	τ 2
N&S	0.05572142	-0,03154683	-0,04809111		1,86079661	
Svensson	0.05570325	-0.03151222	0.02464373	-0.07449321	1.75867062	1.8224406

The parameters can be interpreted as being related to the levels of the long-term interest rate, the short-term interest rate and the slope and hump of the curve. The two additional parameters of the Svensson model can be interpreted as allowing for an additional hump in the curve. Both models are constrained to converge to a constant level and confirm the unbiased expectations hypothesis, whereby agents perceive similar and constant expectations at the long end. From Table 6 above, the parameter values of the two models are very similar, with only marginal differences at the third decimal place. Similar results and trends can be found for other days.

Spline-based models have a much larger number of parameters that should allow for a better fit and flexibility of the curve, which is determined by the number and location of the knot points and the settings of the roughness penalty function[33]. The relatively large number of parameters, however, makes it difficult to demonstrate, compare and interpret the parameter values in an economic sense.

Table 7 below shows the in-sample test results (covering the whole period of two years of data). The table presents the weighted mean absolute errors (WMAE), the root mean squared error (RMSE) and the hit rates[34] for the full dataset, broken down by maturity bracket for the four models together.

[33] One can vary the level and maturity intervals of the roughness penalty or the smoothing parameters, L, μ and S determining the variable roughness penalty function.

[34] Please see Appendix 1 for the statistical description of the various error measures.

Table 7
Summary of the in-sample test results

Summary statistics		Period: **1/4/2003-31/3/2005**								
		Maturities								
Models	Statistics	All	0.25 -1	]1-2	]2-3	]3-5	]5-7	]7-10	]10-20	]20-30
Nelson & Siegel (Central bank of Greece)	WMAE	0.04100	0.03600	0.03400	0.02300	0.03000	0.03700	0.05400	0.06800	0.06600
	RMSE	0.05700	0.05400	0.04400	0.03100	0.04100	0.04600	0.06800	0.08300	0.08800
	Hit rate	0.15017	0.21715	0.21894	0.17890	0.15930	0.07561	0.08181	0.04722	0.13022
Svensson (Central bank of Greece)	WMAE	0.03700	0.03200	0.03400	0.02300	0.02900	0.03700	0.05000	0.05900	0.04900
	RMSE	0.05000	0.04700	0.04300	0.03000	0.03900	0.04400	0.06100	0.07000	0.06700
	Hit rate	0.15591	0.23149	0.22408	0.19176	0.15174	0.06482	0.07031	0.05851	0.17500
Nelson & Siegel (Deutsche Bundesbank)	WMAE	0.05330	0.05970	0.05300	0.03750	0.02870	0.03040	0.03940	0.05800	0.06540
	RMSE	0.07100	0.10280	0.06050	0.04280	0.03560	0.03890	0.05500	0.07610	0.08760
	Hit rate	0.11350	0.17240	0.09830	0.10520	0.11520	0.09760	0.09600	0.06530	0.08580
Svensson (Deutsche Bundesbank)	WMAE	0.04120	0.04500	0.03580	0.02570	0.02880	0.03770	0.04440	0.06150	0.06470
	RMSE	0.06620	0.09110	0.04650	0.03270	0.03640	0.04240	0.05370	0.07820	0.09160
	Hit rate	0.11940	0.18100	0.15860	0.19370	0.12300	0.03510	0.04160	0.05470	0.11770
Nelson & Siegel/Svensson (Banque de France)	WMAE	0.01752	0.00205	0.01091	0.01829	0.01608	0.01412	0.02124	0.00889	0.00081
	RMSE	0.02488	0.00224	0.01253	0.02423	0.02533	0.01894	0.04086	0.02006	0.00102
	Hit rate	0.39690	0.92500	0.23810	0.28616	0.21716	0.16592	0.10282	0.15752	0.31416
Waggoner with step-wise linear penalty function (Central Bank of Greece)	WMAE	0.04500	0.04200	0.03400	0.02300	0.03400	0.03900	0.05800	0.07300	0.07100
	RMSE	0.06100	0.06000	0.04300	0.03000	0.04500	0.04800	0.07400	0.08900	0.09200
	Hitrate	0.13065	0.18675	0.19267	0.15564	0.12903	0.06502	0.06872	0.03778	0.10678
Waggoner with continuous penalty function (Bank of England)	WMAE	0.03326	0.07862	0.02621	0.03231	0.02339	0.01675	0.01502	0.02281	0.01718
	RMSE	0.08971	0.10676	0.04815	0.11298	0.09360	0.06272	0.06021	0.08906	0.07654
	Hitrate	0.29434	0.24929	0.40343	0.37590	0.42808	0.28288	0.32379	0.05725	0.04067

Table 7 demonstrates that each of the four models produces a significantly high in-sample fit with a particularly low weighted mean absolute error.

For the parsimonious models, the error term (WMAE) is within the range of [1.8 to 5.3] basis points for all maturities. The root mean squared error statistics are larger than the weighted average price errors per construction and show similar indications and results, when comparing the two parsimonious models. The lowest error terms can be found for the hybrid model as tested by the Banque de France, which is explained by the switching between the two models, depending on the parameter value for the Nelson & Siegel model. Generally speaking, it can be seen that the Nelson & Siegel model has an insignificantly higher weighted mean absolute error term of [4.1 to 5.3] basis points than the Svensson model ([3.7 to 4.1] basis points), taking all maturities together.

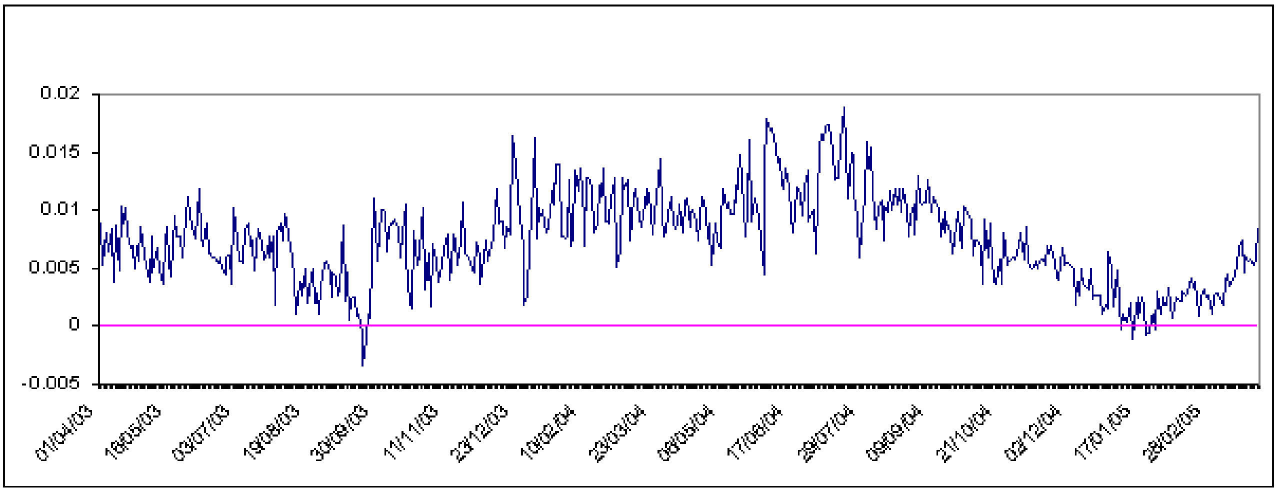
If one compares the performance of the models across the short-, medium- and long-term maturity brackets, the statistical results indicate that the Nelson & Siegel and Svensson models fit the dataset significantly well across all maturity brackets with a slightly better fit of the Svensson model for the long-term maturity bracket of above ten years. The error statistics seem to be insignificantly lower in the medium-term maturity bracket and increasing in the long-term maturity bracket. Within the short- and medium-term brackets, the WMAE statistics are within the range of [2.3 to 5.9] and [2.3 to 4.5] basis points for the Nelson & Siegel model and the Svensson model, respectively. These low values indicate very high goodness-of-fit statistics across all maturities. A similar tendency is seen for the RMSE statistics.

Similarly, both spline-based models produce very high in-sample goodness-of-fit statistics, where the weighted mean absolute errors are considerably low and within the range of [3.3 to 4.5] basis points, and the root mean squared error statistics are within the range of [6.1 to 9.0] basis points for the Waggoner and VRP models respectively. The results of the spline-based models show a very similar picture, where the WMAE statistics decrease in the medium-term maturity bracket to 2.3 basis points and then increase again in the long-term maturity bracket with error statistics of up to seven basis points for the Waggoner model. The VRP model has very low error statistics for the long-term maturity bracket within the range of a couple of basis points, whereas the errors are up to seven basis points for the short-term maturity bracket. The RMSE statistics show a similar tendency. Again the spline-based models demonstrate very high goodness-of-fit statistics.

A visual presentation of error statistics makes it easier to read. One way of comparing the error statistics visually is to subtract the daily error terms between the two models, for instance the results of the Nelson & Siegel and Svensson models[35] as shown in Chart 15. A positive value indicates that the Svensson model provides lower error statistics than the Nelson & Siegel model for a given day and a negative value indicates that the Nelson & Siegel model performs better.

Chart 15

Visual presentation of the daily spread of the residual mean squared errors between the Nelson & Siegel and Svensson models



In the large majority of cases, the spreads are positive, indicating that the daily root mean squared errors resulting from the Svensson modellisation provide slightly lower error statistics than the N&S estimation over the two-year period, and,

[35] The spread was calculated each day (t) as follows: Spread (t) = RMSE [N&S] (t) – RMSE [Svensson] (t).

consequently, indicates that the Svensson model is preferable, with the caveat that the spread values are significantly low and therefore not conclusive.

Turning to the hit-rate statistics, since the error terms are all significantly low, a very low threshold level is set for assessing the hit rates. This means that if the data point is within plus/minus 1.5 basis points from the estimated curve, the observation will be counted as “hitting the target”. The hit rate therefore represents the percentage of observed yields that lie within 1.5 basis points from the estimated curve. Models with a high percentage of hits fit the dataset better than a model with a relatively lower percentage of hits.

The statistics indicate, overall, that 11-16% of all observed yields are within 1.5 basis points from the estimated curve for the two parsimonious models. In the special case of the hybrid model, 40% of the observations are within the narrowly defined confidence interval. The hybrid model performs particularly well in the maturity bracket of three months to one year, with a hit rate of over 90%, and the hit rates also remain above 10% for long maturities.

For the two spline-based models, the hit statistics indicate that approximately 13% and 29% of all observations are within the confidence interval of the estimated curve for the Waggoner and VRP models, respectively. The overall hit rates for the VRP model are very high for the full period and significantly higher than the Waggoner model; however, for the long-term segments (above ten years), the hit-rate statistics significantly decrease to 4-5%, caused by the greater smoothing after the ten-year maturity.

A priori these hit statistics may not seem to reach high levels; however, this is related to the extremely tight confidence interval applied, where the observed data points may not deviate by more than 1.5 basis points on either side of the estimated curve. If, for instance, the confidence interval is increased to three or five basis points, the hit-rate statistics reach levels of 72% and 98% for the Svensson model. A hit-rate statistic of 95% is remarkably high and demonstrates a very good fit.

6.2 Flexibility & goodness-of-fit criterion (out-of-sample results)

The out-of-sample test results are important as they may be more indicative of the underlying term structure than the in-sample testing; this is because this test provides an indication of whether the estimated curve can accurately price a bond which has not been used to estimate the actual curve.

The out-of-sample test results covering the whole period of two years of data are shown in the following table.

Table 8

Summary of the out-of-sample test results

Summary statistics		Period: **1/4/2003-31/3/2005**								
Models	Statistics	Maturities								
		All	0.25 -1	]1-2	]2-3	]3-5	]5-7	]7-10	]10-20	]20-30
Nelson & Siegel (Central bank of Greece)	WMAE	0.04100	0.03700	0.03600	0.02500	0.03000	0.03700	0.05400	0.06600	0.06100
	RMSE	0.05600	0.05500	0.04700	0.03300	0.04000	0.04500	0.06800	0.08000	0.08400
	Hit rate	0.14714	0.20543	0.21030	0.18465	0.15305	0.07235	0.08512	0.04114	0.11867
Svensson (Central bank of Greece)	WMAE	0.03700	0.03300	0.03500	0.02500	0.02900	0.03700	0.05000	0.05900	0.04400
	RMSE	0.04900	0.04800	0.04500	0.03300	0.03800	0.04400	0.06100	0.06900	0.06300
	Hit rate	0.14920	0.21177	0.21289	0.19757	0.14927	0.06088	0.07438	0.04517	0.15156
Nelson & Siegel (Deutsche Bundesbank)	WMAE	0.05390	0.06070	0.05340	0.03670	0.02890	0.02950	0.03870	0.05540	0.06000
	RMSE	0.06410	0.07690	0.05750	0.03990	0.03390	0.03390	0.04840	0.06190	0.06840
	Hit rate	0.11460	0.18370	0.09250	0.10460	0.10840	0.10220	0.08830	0.05100	0.08480
Svensson (Deutsche Bundesbank)	WMAE	0.04230	0.04740	0.03630	0.02600	0.02880	0.03760	0.04350	0.05860	0.05930
	RMSE	0.05890	0.06580	0.04170	0.03040	0.03430	0.04050	0.04960	0.06400	0.06850
	Hit rate	0.12070	0.16680	0.15300	0.21100	0.12490	0.03360	0.03850	0.05130	0.14060
Nelson & Siegel/Svensson (Banque de France)	WMAE	0.03398	0.02548	0.00586	0.00366	0.01126	0.01175	0.01270	0.01671	0.00354
	RMSE	0.07214	0.02914	0.00796	0.00466	0.01283	0.01351	0.01557	0.03598	0.00354
	Hit rate	0.34811	0.57778	0.66667	0.50000	0.00000	0.15897	0.20313	0.16705	0.00000
. Waggoner with step-wise linear penalty function (Central Bank of Greece)	WMAE	0.03900	0.03700	0.03500	0.02500	0.03200	0.03800	0.05200	0.06100	0.04400
	RMSE	0.05100	0.05100	0.04500	0.03200	0.04200	0.04600	0.06500	0.07100	0.06000
	Hitrate	0.13243	0.18078	0.18506	0.15695	0.12550	0.06294	0.07491	0.03703	0.10206
Waggoner with continuous penalty function (Bank of England)	WMAE	0.03569	0.08038	0.02694	0.03110	0.02193	0.01238	0.01928	0.01662	0.00251
	RMSE	0.08617	0.08070	0.04137	0.07003	0.06577	0.03677	0.06794	0.06427	0.01154
	Hitrate	0.30026	0.22296	0.36013	0.32786	0.40803	0.30528	0.20932	0.02861	0.00697

The first indication from the above table is that the various error term measures remain significantly low for the full two-year period. In addition, the results are very similar to the in-sample test results. The second indication that can be drawn from the table is that there is a slight increase in the error terms per increasing maturity brackets, which was also observed using the in-sample test. Nevertheless, the error terms remain at very low levels, in absolute terms, with values between [3.4 and 5.4] basis points for the WMAE statistics and between [5.0 and 7.2] for the RMSE statistics, with slightly increasing though insignificant error terms for the longer maturity bracket of 20-30 years, except for the hybrid model.

For the spline-based models, the table indicates that for the fitting errors, broken down by maturity interval, the difference in the WMAE statistics between the two models is smaller for the short maturity brackets and larger for the longer maturities (this is similar and even more pronounced for the in-sample results). This might be a consequence of the selected penalty function which imposes a substantially higher penalty on changes in the slope of the curve at longer maturities, or it might also be a result of the weighting.

During the full two years of data, the in-sample and out-of-sample fitting errors are higher for the Waggoner model than for the VRP model. The analysis of the different sub-periods confirms similar findings; the fitting error of the VRP is never, or in very few cases, higher than that of the Waggoner model.

Overall, the out-of-sample test results demonstrate significantly low weighted mean absolute errors and root mean squared errors, with similar hit-rate statistics to those for the in-sample test results.

Despite the use of statistical tests for evaluating models, a visual inspection is indispensable. The visual inspection helps you to judge if the statistical tests are reasonable and also if the curves behave as expected. For instance, one test result is that the hit rates of the spline-based models are relatively low for maturities above ten years. This is similar to the parsimonious models and therefore not surprising as such. However, using a visual inspection, one realises that the fit of the spline-based models for maturities above ten years behaves in a highly unexpected manner and moves rather irregularly at long maturities. This can be visually inspected by viewing the daily estimations over the two-year period and converting these daily results into a movie visualising the performance of the yield curve estimations over the two-year period (see Appendix 5 for the respective videos). Comparing, for instance, the Svensson video with the Waggoner video, the asymmetric movements of the spline-based models for long maturities above ten years can be seen (see the two snapshot examples in Chart 16 below).

Chart 16

Example of a visual inspection of the spline-based model for maturities above ten years

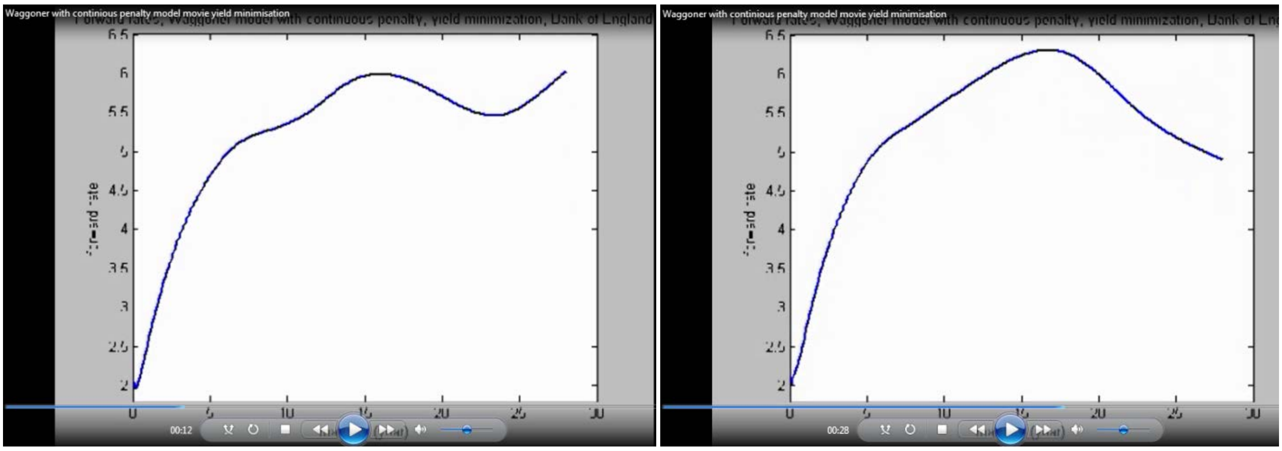


Chart 16 demonstrates the importance of visual inspections. The spline-based model indicates a yield of approximately 6% at the 30-year maturity (left chart) while the curve (right chart) indicates a yield of approximately 5% at the 30-year maturity and with totally opposite directions. While these two curves are at different points in time, the underlying high-quality datasets do not support those sporadic movements at the long end of the curve.

The hit-rate statistics are low for long-term maturities (generally for all estimations) and express a value showing how many observations there are within the defined interval band of the curve and not how far away they are from the interval band. Therefore, the visual inspections are important tools to observe abnormal behaviours and to confirm the statistical results.

Based on the flexibility & goodness-of-fit (in-sample and out-of-sample) statistics, it can be concluded that all four models perform well and provide very low error statistics in combination with high hit-rate statistics. The low error statistics demonstrated by the two parsimonious models during the sample testing periods and for the maturity brackets indicate a slight favouring of the Svensson model, which seems to consistently provide better in-sample and out-of-sample statistical results than the Nelson & Siegel model, although it only demonstrates insignificantly better goodness-of-fit and hit-rate statistics.

If the overall performance of the two models during the individual selected time periods (the period of an expected interest rate decrease and the period of high market volatility) is compared, both models perform well and continue to obtain high goodness-of-fit statistics. As confirmed within the literature review, the Svensson model is more sensitive to changes in interest rates and, therefore, any interest rate change may have a larger effect in Svensson estimations than in Nelson & Siegel estimations. However, using the selected sample period, both models seem to perform similarly and the Svensson model does not demonstrate a superior fit. This could, however, be explained by the relative stability and smoothness of interest rates during the two-year data sample period. It should be noted that the Svensson model is used by the Banque de France in 405 out of the 502 daily estimations as part of the hybrid model, indicating that the Svensson model is selected in more than 80% of the daily curve estimations.

The goodness-of-fit test results demonstrate that all four models capture the movements in the underlying term structure very well and the estimated curves accurately price a bond which has not been used to estimate the curve.

6.3 Robustness test

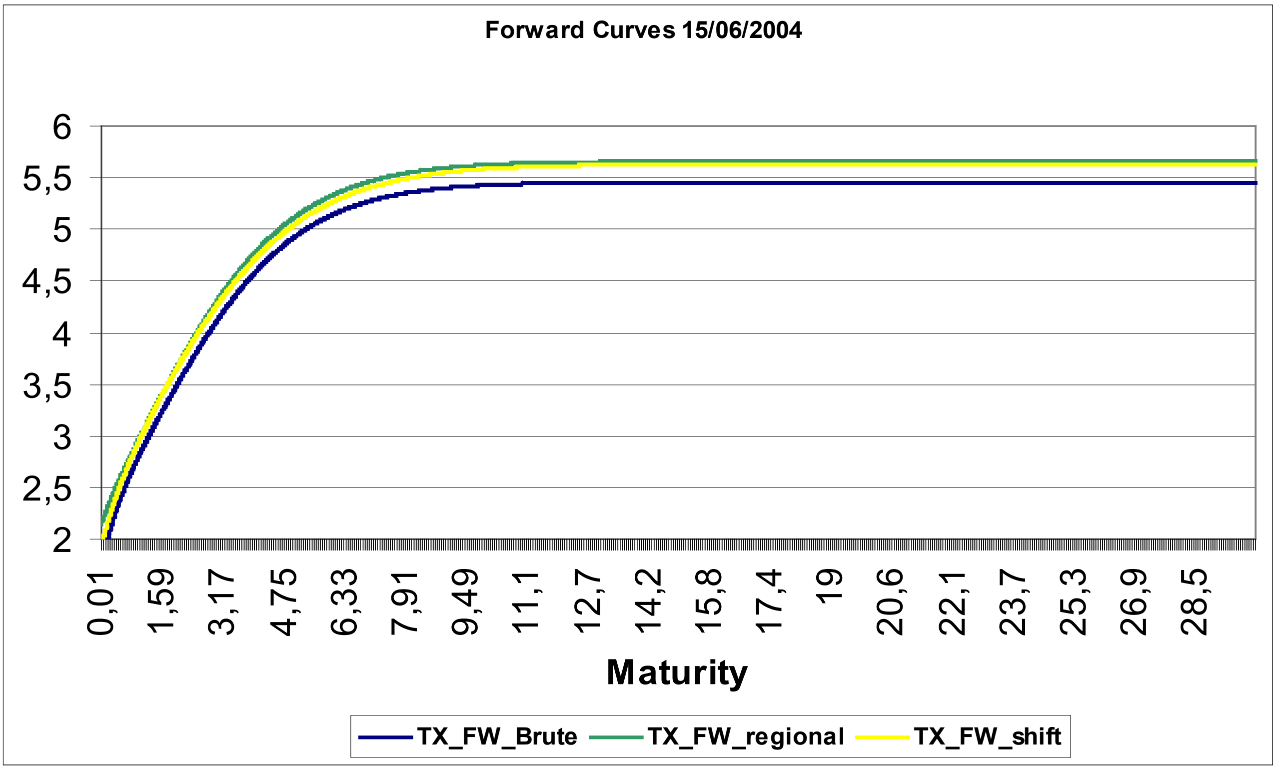
The robustness test criterion measures the sensitivity of the estimated model to price changes and provides a way of ensuring that changes in the data at one maturity (such as at the long end) do not have a disproportionate effect on other maturities (such as at the short end).

As an illustration of the robustness test, the sensitivity of the hybrid model is demonstrated by performing three different shifts of the underlying bond data and checking the effect on the estimated models (15 June 2004):

- one global shift of 15 bps for the entire curve (labelled "brute");
- one local shift of three standard deviations for one chosen bond with a residual maturity of one year (labelled "regional");
- one regional shift of 30 bps for bonds with a residual maturity of less than one year and a shift of 20 bps for bonds with a residual maturity of more than one year (labelled "shift").

Chart 17

The impact of the robustness test as demonstrated by shifting prices for the hybrid model



The graph above shows the resulting yield curves after three shifts were applied. The estimation method is robust if the effect of the local changes on the resulting yield curve is small and the effect of a global change is proportionate. The results seem to suggest that the models are reasonably robust: they tend to reflect closely global changes, while remaining robust following local perturbations.

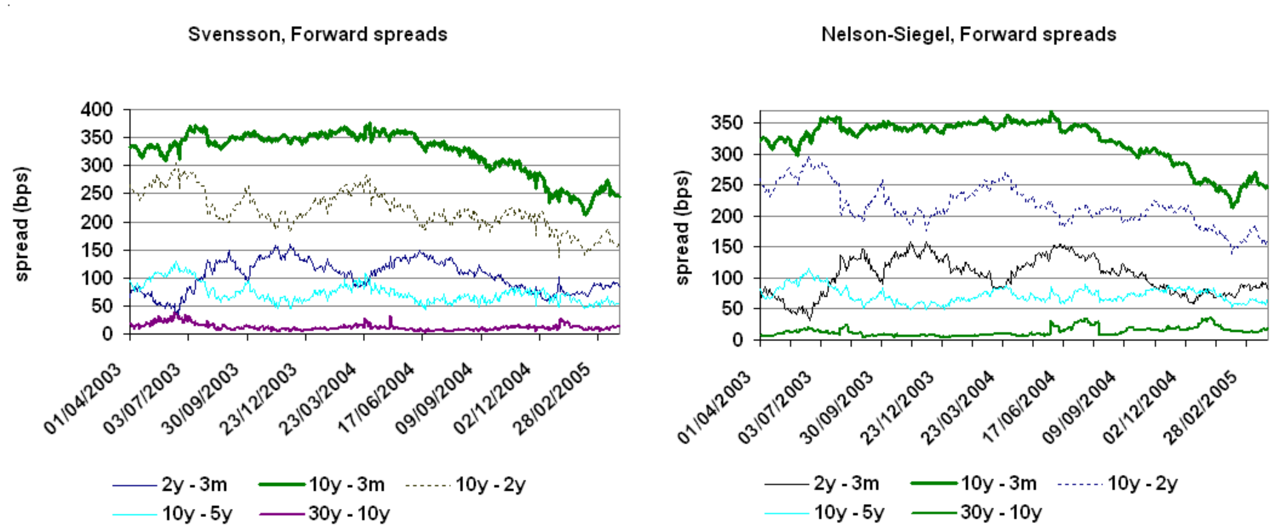
6.4 Comparing smoothness

The smoothness criterion aims to provide a recommendation as to which models generate relatively smooth curves. The changes in the slope can be studied by calculating the spreads between different maturity levels. Indeed, the variations in the different spreads indicate changes in the slope and thus allow us to determine the changes in the curve over time.

Chart 18 shows, as an example, the spread of the forward rates using the Svensson model and the Nelson & Siegel model, respectively, whereas Chart 19 provides an example for the VRP and Waggoner models.

Chart 18

Spread between different maturities using the Svensson and Nelson & Siegel models



Both the Svensson and Nelson & Siegel curves behave similarly for the short- and medium-term maturity spreads. Within the time period under review, the changes in the curve's slope seem relatively smooth ([ten years – two years] and [ten years – three months]) and the spreads indicate a flattering curve during the second half of 2004. The difference in volatility between both models is more obvious for the long maturity spreads ([30 years – ten years] and [ten years – three months]). For both models, spreads are more volatile when the shortest three-month maturity is included.

Chart 19

Spreads between different maturities using the Waggoner and VRP models

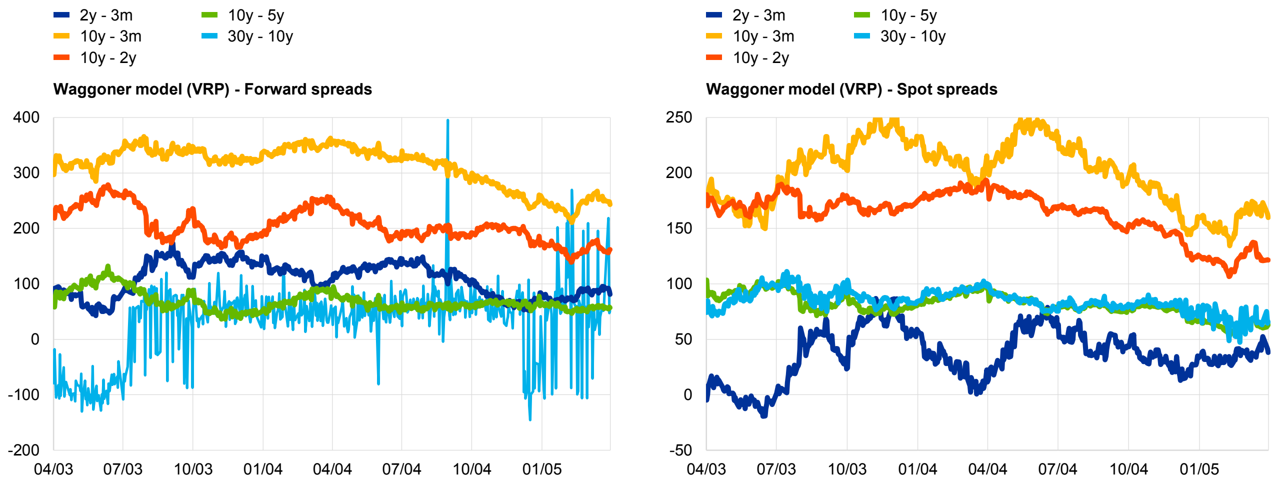


Chart 19 shows the changes in the forward rates and spot spread over time using the VRP and Waggoner models respectively. Both graphs indicate that the curves behave similarly at the short- and medium-term maturities. Note that the curve of the VRP model becomes significantly more volatile than that of the Waggoner model if the 30-year yield is included, in particular during the latter part of the sample period. While it starts relatively smoothly, a few abrupt breaks can be observed throughout the time periods and in particular during the latter part of the sample. These abnormal behaviours cannot be explained by changes in the slope of the yield curve and indicate that the model may not be smooth between the maturities of ten years and 30 years.

This finding is not intuitively expected, nor is it confirmed in similar studies. One possible explanation might be that the maximum maturity estimation varied over time and was often below 30 years.

Within the time period under review, the changes in the curve's slope seem relatively smooth for the [ten years – two years] and [ten years – three months] spreads, and also indicate a flattening curve during the second half of 2004.

6.5 Summary of test results

From the above test results, it can be concluded that all four models (the two parsimonious and two spline-based models) can represent the term structure of euro area interest rates, as they all demonstrate highly consistent statistical results across the maturity spectrum, which are better than normally found in the literature. This result is based on a large sample size that tests the performance of the models independently during several different periods using the two years of available daily data. Two of the main reasons for these very good statistical results may be the good quality of the filtered bond data used as input to the models and the significantly larger bond sample than is normally available and applied in the calculation of national yield curves. A third reason for these good results might be the relatively stable debt securities markets, where monetary policy interest rates only decreased once during the two-year data period. This could also mean that some of the yield curves might be over-parameterised, as several combinations of parameters describe nearly the same curve. Based on the results of the testing exercise, the hybrid model provided the best goodness-of-fit statistics.

Overall, the results from the Svensson model seem to be slightly better than those of the Nelson & Siegel model. This is confirmed by both the statistics presented above and the literature.

The robustness results suggest that both parsimonious models are reasonably robust since the effect of local changes is small and the effect of a global change is proportionate. While comparing the slope of the curves of different maturities, it was found that the spread is more volatile when the shortest three-month maturity is included, while changes in the slope seem relatively smooth throughout the period under review. Both parametric models show well-represented results in terms of the smoothness.

During the empirical testing, it was found that the two parsimonious models are sensitive to the selection of the starting values and the applied optimisation algorithm, and further research should be encouraged to assess the impact of these two relationships.

The spline-based models are also supported by very good goodness-of-fit statistics for both the in-sample and out-of-sample statistics and the very low error statistics. While comparing the slope of the curves (the spreads between different maturities), the results indicate that the curves are relatively smooth for short-to-medium-term spreads, whereas when longer maturities (10-30 years) are included, the slope becomes volatile, significantly more for the VRP model than for the Waggoner model.

During the assessment of the testing exercise, it was identified that the comparison of the spline-based models is more sensitive to model-based factors than the parsimonious models. In particular, the results may differ depending on: (1) the applied optimisation algorithm; (2) the optimisation of the smoothing parameters; (3) the selection of the penalty function; and (4) the setting of knot points. There is no conceptual reason why either variant should be superior; rather, it is an empirical matter where different model-based factors might have an influence. An important consideration is the trade-off between the smoothness and the flexibility of the curves.

This trade-off was studied when optimising the smoothing parameters. A combination of smoothing parameters that is too flexible would reduce the fitting errors but estimate unsmooth curves and vice versa. Therefore, different combinations of smoothing parameters were tested to try to find the optimal combination using the same dataset[36]. Furthermore, the number and location of the knot points also play a role as the optimal smoothing penalty is conditional on a given set of knots. Once the number of knot points is set, an optimal trade-off between flexibility and smoothness can be achieved by changing the variable roughness penalty by varying the:

- level and maturity intervals of the roughness penalty for the Waggoner model; and
- smoothing parameters, L, μ and S, determining the variable roughness penalty function[37].

Furthermore, the Waggoner model is likely to be more stable in response to a small change in bond prices on account of the "minimum support" feature of spline models.

A comparison of the spline-based methods, in particular those based on the VRP model, appears to be more complex than that of parametric models. The main

[36] The combination L=100,000 ; S=0.1 ; μ=1 was chosen as the best-performing combination in terms of smoothness and flexibility and therefore all of the test results for the VRP model are based on this combination of smoothing parameters.

[37] Anderson and Sleath defined their penalty function λ as: λ(m) = 10 ^ [log10 (L) – (log10 (L) – log10 (S)). exp(-m/μ)] where L, S and μ are three smoothing parameters to be optimally fixed before estimation. m is the maturity.

difference is that the balance between smoothness and flexibility is deliberately chosen by the modeller in line with his/her objectives and the nature of the underlying bond data. For example, the choice of a step-up penalty function might be based on the consideration that the euro area bond market might be effectively segmented into a short-term bond market and a long-term bond market. This assumption is reasonable and is also current practice within euro area debt securities statistics[38].

The spline-based test statistics may also vary depending on the number and location of the knot points and on the parameters of the penalty function, which are the result of an ad hoc optimisation procedure. In addition, these choices directly affect the robustness of the resulting curve itself. The complexity of the optimisation process and the high number of parameters involved might actually imply a lower degree of transparency of the spline-based methods compared with the parametric ones. On the other hand, the attractiveness of the spline-based methods partly lies in the empirical nature of the estimation procedure and its extreme flexibility. These results call for further research to properly understand the trade-offs and interdependencies involved in calibrating the smoothing parameters and the penalty function and in setting the knot points and the applied optimisation algorithm.

[38] The euro area debt securities statistics are split into “short-term debt securities” (residual maturity of one year or less) and “long-term debt securities” (with a residual maturity of more than one year). However, within national debt securities markets, “short-term debt securities” may also include debt securities with a residual maturity of up to 18 months.

Conclusions

Bond prices reflect market participants' views on interest rate levels in a forward-looking way. The level of market interest rates typically depends, among many other factors, on the residual maturity of the underlying bonds. The relationship between interest rates and the residual maturity is referred to as the term structure of interest rates and can be displayed by the yield curve. The yield curve offers a particularly extensive and useful set of information about the expected path of future short-term rates and the outlook for economic activity and inflation, which is valuable for monetary policy purposes. Therefore, yield curves – via the calculation of implied forward rates – contain information on market participants' expectations of the relative level of future short- and long-term interest rates at a certain point in time as expressed by the slope of the yield curve. The slope of the yield curve can be used as an indicator of the outlook for economic activity. A steepening of the curve often anticipates an acceleration of economic activity, while a flattening of the curve often indicates a slowdown of economic activity. The explanation for this is that in normally functioning financial markets, a large positive spread between long- and short-term interest rates may indicate that the market anticipates an increase in short-term interest rates because of a more positive outlook for economic growth. If the yield curve is decomposed into a real interest rate and inflation component, the longer end of the yield curve may also reflect market participants' views about trend developments in inflation. The real interest rate component can be approximated by the yields on inflation-linked government bonds, although there are relatively few government bonds linked to the consumer price index. The inflation component can be created by subtracting the yield curve from a comparable inflation-linked yield curve ("break-even inflation rates"), which can be used to approximate developments in inflation.

Yield curves are used by the central banking and financial community across the world as a leading financial market indicator. Almost all European Union national central banks produce (and some release) national yield curves. However, following the creation of the Monetary Union, there has been a complementary need for calculating meaningful and representative euro area financial market indicators, including various representations of euro area yield curves for monetary policy purposes and for comparing the euro area as a whole with other economic areas.

The first objective of this paper is to supplement the comprehensive albeit mainly theoretical literature on yield curves by dealing with the more operational side of term structure estimations and to provide a conceptual framework for selecting a high-quality bond sample before starting any testing. The focus on selecting high-quality bond data is important, as financial market indicators provide leading and complementary information in any decision-making process, and therefore nothing is more important for the reliability and predictability of financial market indicators than using good quality data. This is often neglected, but it is a crucial step before starting any model estimations and testing exercise.

The second objective of this paper is to present and test four types of yield curve models and demonstrate their performance using several statistical tests and examples. These tests were undertaken together with experts from euro area national central banks. The result of this testing exercise has paved the way for a prudent recommendation, implementation and subsequent release of daily euro area yield curves on the ECB's website.[39]

With regard to the sample population, it is vital that practitioners obtain a comprehensive overview of the necessary descriptive statistics of the dataset as part of cleaning, outlier removal and selecting the final data population for the empirical yield curve modelling. Liquidity considerations also play a major role in the selection process as the more liquid a market is, and the more frequent trades are, the more representative the informational content of prices in the market is. An assessment of price quality near the market close is required. Furthermore, the volatility of yields across bond residual maturity classes needs to be analysed, which may lead to the elimination of bonds with a short residual maturity, as this class seems to demonstrate significantly higher volatility than other maturity classes. The descriptive statistics lead to twelve fundamental recommendations on how to derive a clean dataset for yield curve estimation. The testing exercise confirmed that a high-quality financial market data population is essential for any model estimations, thereby providing users with sound and reliable yield curve estimations.

Against this background, the following framework for testing yield curves is established.

Table 9

Conceptual framework for selecting a high-quality bond sample for yield curve estimations

Dimension	Recommendations for selecting a high-quality yield curve sample for implementing yield curves
Supplier of data	Knowing your data source. There is a clear need to assess and compare the representativeness and quality of your data source. An analysis should be conducted of which commercial and institutional data vendors can deliver the relevant high-quality bond data from regulated and non-regulated trading places. This should include a comparison of which data vendors can comply with the following operational criteria, with the objective to select a reliable and sustainable data supplier: (a) which fields are mandatory and which fields are available; (b) methodology and definitions of fields (including price – see separate category below) and quality of field values; (c) ability to provide structured and stable data formats and apply international standards; (d) data submission methods and channels; (e) ability to provide test files and flexibility in updating repetitive errors and to be receptive to enhancing quality and price information; (f) price and commercial terms; (g) any confidentiality restrictions for storing and potentially publishing data and results.
Type of instrument	There is a clear need to be able to identify the characteristics of the bond sample provided. Descriptive and distribution statistics are used to compare the differences between markets, instruments and maturities and to select representative and homogeneous debt securities instruments. Should a yield curve be constructed that combines debt securities and money market instruments for the short-term maturity? It is recommended to construct separate curves according to a consistent set of financial instruments and not to mix financial instruments; debt securities and money market instruments have different statistical properties and react differently to monetary policies and shocks, in particular for longer maturities. If money market instruments are considered, it is recommended to use EURIBOR rates, which exhibit the most similar characteristics vis-à-vis government bonds. Statistical analysis and visual inspections are needed to make decisions about including/excluding

39 Euro area yield curves.

Dimension	Recommendations for selecting a high-quality yield curve sample for implementing yield curves
	bonds with certain characteristics and special features (such as Brady bonds, convertible bonds and bonds with embedded options) and at the same time to establish rules for possibly including bonds with fixed or variable rates, zero-coupon bonds, inflation-linked bonds, perpetual bonds or “strips”. This is all part of knowing your sample and its properties. This also has consequences for calculating and deriving yields.
Sector	A conceptually sound yield curve represents an unambiguous sector classification, comparable with international standards. In this case, it must be feasible to identify and extract the “central government” sector according to international standards such as the European System of National and Regional Accounts (ESA 2010) sector classification.
Credit risks	Bond prices are influenced by, inter alia, credit risks. Credit risks can be approximated by using credit ratings from rating agencies at individual bond level and/or sector level. Credit risks can be assessed using ratings as a selection criterion to show a perceived “risk-free” yield curve using top-rated (AAA) debt securities of sovereign Member States – though many countries may not obtain this top rating from the rating agencies.
Liquidity	Liquidity considerations play a major role in the selection process as the more liquid a market is, and the more frequent trades, the more representative the informational content of prices is. Liquidity can be used as a selection criterion by defining threshold levels for trading volumes and/or bid-ask spreads during a period of subsequent trading. A quantitative threshold of average daily trading volumes over a period of time and a small bid-ask spread level are recommended, although these thresholds need to be adapted to the characteristics and dynamics of the relevant financial markets and trading places.
Price type	An assessment of the price definition, calculations and price quality at a similar reference point in time is essential. Bond prices differ according to how data vendors express prices and can have a significant impact on yield curve estimations. Descriptive price statistics can reveal the difference between using “real traded prices” or “mid-price” quotation close to 17:00 (market close). Other markets may express instruments in “yields”, which then need to be converted to prices, and other platforms may offer binding quotes which are immediately executable. Price information, price calculations and price quality differ significantly depending on the data supplier, trading platform and market. This descriptive step cannot be neglected or even be taken for granted.
Maturity spectrum	The volatility of yields within and across bond classes of residual maturity needs statistical analysis, which could lead to the elimination of bonds with a residual maturity of less than three months, as this class may demonstrate significantly higher volatility than other maturity classes. In this case, the yield curve is estimated from three months to 30 years of residual maturity if there are a sufficient number of bonds represented within each maturity bracket.
Removal of outliers	Despite the application of the selection criteria above, individual bond yields may still deviate significantly from the norm of similar types of bonds and have to be removed. Bonds are removed if their yields deviate, for instance, by more than two times the standard deviation from the average yield of bonds within the same maturity bracket. This procedure can be repeated a second time for removing additional outliers.
Special effects and adjustments	There is a positive relationship between coupon and price and an inverse relationship between price and yield. If the coupon is significantly different across countries, then the yield and its changes may not be fully comparable and therefore coupon-effect adjustments could be applied. Similarly, adjustment should be made for bonds which are associated with special national incentive schemes or national tax regimes. In practice, there is no suitable way of making these direct adjustments; however, it is expected that they would be removed as part of the outlier detection mechanism if the coupon/price deviates from the norm. However, importantly, national bonds are adjusted according to day-count conventions and settlement practice to ensure comparability across national bonds. As euro area bond markets are fairly integrated, specific national public holidays may not necessarily affect the calculations except during TARGET holidays, which are specific days on which financial markets are closed.
In-sample and out-of-sample population	Selecting yield curve models based only on in-sample statistics may be misleading and therefore the performance of models should also be assessed according to an out-of-sample population. In this case, all bonds can be ranked in ascending order of residual maturity and, for instance, one out of ten bonds can be selected for the out-of-sample population. This method provides an out-of-sample population where bonds are represented within all maturity classes and can be used to compare the performance of the yield curve estimations using an out-of-sample population.
Sample size	Applying the various selection criteria above may well lead to a reduced sample size, though with higher quality. Higher quality is always preferred to quantity. A trade-off may be needed in the event of small sample sizes and where few bonds are shown within the maturity spectrum. In this case, using averages or increasing the standard deviations for outlier detection may be possible options.
Time periods	The performance of models needs to be tested separately during different market conditions and expectations, and therefore different time periods need to be selected for testing. Models could be assessed during periods of stable markets, periods with market expectations of interest rate increases, periods of decreasing interest rates, and periods of particularly high bond price volatility.

In addition to the above 12 golden data recommendations, this paper also reviews the literature on yield curves and, based on the extensive experience of national central banks, two parsimonious yield curve models[40] and two spline-based yield curve models[41],[42] were comprehensively tested.

All models underwent the same statistical tests[43] and were evaluated according to their performance using exactly the same bond dataset.

Statistical tests were conducted for each model separately, first using an in-sample population and second an out-of-sample population during different time periods[44] and during the full two years of daily sample data. Flexibility & goodness-of-fit tests, as well as robustness and smoothness tests, were applied to all four models and their performance was assessed.

The conclusion of the testing exercise is very promising. All four models can represent the term structure of interest rates for the euro area, produce smooth yield curves and provide low error statistics and high fit statistics.

An evaluation of the three performance criteria of the parsimonious models shows that the Svensson model performs slightly better than the Nelson & Siegel model in terms of the flexibility & goodness-of-fit statistics and for both the in-sample and out-of-sample tests. This confirms similar results in the literature. Similarly, both parametric models show robust and similar results for the smoothness and robustness of the estimated curves. For the latter test, the effect of local shocks in the data was small and the effect of global changes proportionate, thus leading to a reliable representation of the term structure of interest rates. The spreads between maturity classes become more volatile if yields of bonds with the shortest (three months) and longest (30 years) maturities are included, while the trend of the slope over time seems relatively smooth. From a more technical perspective, it was identified during the analysis phase that the results of the parsimonious models are sensitive to the initial values of the starting parameters and the selected optimisation algorithm. It can therefore be recommended that the previous available parameter values are used as starting values for the algorithm.

Similarly, each of the two spline-based models produce significant in-sample and out-of-sample goodness-of-fit statistics, where the error statistics are considerably low and increasing (as expected) for longer maturities. Using the spreads between different maturities, the results indicate that the slopes of the curves are also rather smooth for short-to-medium-term spreads, but become quite volatile if the 30-year government bond yield is included in the VRP model. Overall, for the whole two-year period, the Waggoner model seems to produce a slightly lower fit than the VRP for maturities up to ten years, although the available evidence does not seem

40 Nelson & Siegel and Svensson model (extended version of the Nelson & Siegel model).

41 Waggoner model with a step-wise penalty function and a continuous penalty function (VPR).

42 Please refer to Appendix 4 for the model properties, objective functions, constraints, starting values, penalty functions and software application used in this testing exercise.

43 Flexibility & goodness-of-fit, robustness and smoothness tests.

44 The performance of each model was tested covering separate periods. Periods were selected that included either a period of strong interest rate cut expectations, a period where the market expected an increase in interest rates, a period of volatility in the debt markets or a period of economic and political uncertainties.

conclusive. The reason for this is that both spline-based models are sensitive to: (1) the applied optimisation algorithm; (2) the fixing of smoothing parameters; (3) the selection of the penalty function; and (4) the setting of knot points. There is no conceptual reason why either variant should be superior; it is rather an empirical matter where these different factors might have an influence. In this respect, it is very important to optimise the smoothing parameters to enable the optimum level to be found in the trade-off between flexibility and smoothness. Against this background, different combinations of smoothing parameters were applied to find and use the optimal combination for the testing exercise. It became evident that the number and location of the knot points play a role, as the optimal smoothing penalty is conditional on a given set of knots.

It can also be concluded that the comparison of the spline-based methods was more complex than that of the parametric models. The main difference is that the balance between smoothness and flexibility is deliberately chosen by the practitioner according to his/her objectives and the nature of the underlying data. For example, the choice of a step-wise penalty function could be based on the consideration – which should be properly documented – that the euro area market might be effectively segmented (short-term versus long-term securities). The complexity of the optimisation process and the high number of parameters involved might actually imply a lower degree of transparency and communication difficulties compared with the parametric ones. On the other hand, the attractiveness of the spline-based methods partly lies in the empirical nature of the estimation procedure and its flexibility. However, further research is needed to fully understand and quantify the effects on the estimations of changing the smoothing parameters and knot points before such yield curves could be produced using spline-based methods.

The very convincing results and performance of all four models may partly be due to the very good quality of the underlying filtered data and the large sample of data used, which is significantly larger than the one normally used for estimating national yield curves. The good results might also be explained by the relatively stable period covered, when monetary policy interest rates were kept stable, with one interest rate decrease during the two-year sample period combined with relatively stable bond markets.

Against this background, it can be concluded that despite the slightly better (though insignificant) statistical performance of the spline-based models, the Svensson model may be the better choice overall for producing yield curves. The model and its results are more transparent and its explanatory power better serves the communication and interpretation needs of financial markets and professional users. It is also suggested that the Waggoner and VRP models be used for comparing yield curves with other major economic areas, although further research is called for to obtain sufficient knowledge of the interdependencies and trade-offs involved in calibrating the smoothing parameters, the penalty function and the selection of knot points and to test the performance of the models at the zero lower bound.

Until then, the ECB publishes two credit risk yield curves using the Svensson model each day at 12:00 CET, making it possible for financial agents to interactively track and download the euro area yield curves[45].

[45] Euro area yield curves.

Appendices

A Appendix 1: Error statistics formulae

There are many measures of the goodness of fit of a model. In this particular case, three statistical errors and one accuracy measure can be used; the formula is for each single day (k), with N observed bonds:

$$RMSE\ (root\ mean\ squared\ errors) = \sqrt{\frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|^2}{N}}$$

where $y_{i,k}$ is the actual yield and $\hat{y}_{i,k}$ the estimated yield on day k.

$$MAE_k = \frac{\sum_{i=1}^{N} \left| y_{i,k} - \hat{y}_{i,k} \right|}{N}$$

where $y_{i,k}$ is the actual yield and $\hat{y}_{i,k}$ the estimated yield

$$WMAE_k = \frac{\sum_{i=1}^{N} \frac{\left| y_{i,k} - \hat{y}_{i,k} \right|}{\sqrt{D_{i,k}}}}{N}$$

where $D_{i,k}$ is the duration of the bond on day k.

$$Hit\ rate\ k = \frac{number\ of\ bonds\ day\ k\ for\ which\ \left| y_{i,k} - \hat{y}_{i,k} \right| \leq 0.03}{N}$$

B Appendix 2: Functional forms of yield curve models

B.1 Parsimonious models

The Nelson & Siegel model specifies a functional form for the instantaneous forward rate, f(t), as follows:

$$f(t) = \beta_0 + \beta_1 exp\left(-\frac{t}{\tau}\right) + \beta_2 \frac{t}{\tau} exp\left(-\frac{t}{\tau}\right)$$

The Svensson model specifies a functional form for the instantaneous forward rate, f(t), as follows:

$$f(t) = \beta_0 + \beta_1 exp\left(-\frac{t}{\tau_1}\right) + \beta_2 \frac{t}{\tau_1} exp\left(-\frac{t}{\tau_1}\right) + \beta_3 \frac{t}{\tau_2} exp\left(-\frac{t}{\tau_2}\right)$$

B.2 Spline-based models

The cubic spline model is computed from an optimised linear combination of the basis spline. The basis can be generated with the De Boor algorithm as follows:

As advised by De Boor, it is convenient to create an augmented set of knot points,

$\{d_k\}_{k=1}^{K+6}$ where $d_1 = d_2 = d_3 = d_4 = s_1, d_{K+4} = d_{K+5} = d_{K+6} = s_K$ and basically, $d_{k+3} = s_k \forall\ k$ in $[1; K]$.

A cubic spline is then a vector of h = P + 2 cubic B-splines defined over the domain. A B-spline is defined by the following recursion, where r = 4 for a cubic B-spline and

$1 \leq k \leq P$:

$$\Theta_k^r(m) = \frac{\Theta_k^{r-1}(m) \times (m - d_k)}{d_{k+r-1} - d_k} - \frac{\Theta_{k+1}^{r-1}(m) \times (m - d_{k+r})}{d_{k+r} - d_{k+1}}$$

for $m \in [0, M]$, with

$$\Theta_k^1(m) = \begin{cases} 1, \forall m \in [d_k; d_{k+1}[\\ 0, \quad\quad else \end{cases}$$

So finally the vector $\Theta^r(m) = (\Theta_1^r(m), \dots, \Theta_P^r(m)) = (\Theta_1(m), \dots, \Theta_P(m))$ is obtained.

Then, any cubic spline can be constructed from the linear combination of this basis:

$$\beta = (\beta_1, \dots, \beta_P)^T$$

B.3 Roughness penalty functional forms

B.3.1 Step-wise variable penalty function

For a given maturity interval $[m_{min}; m_{max}]$, Waggoner defined a step-wise penalty function, constant across three maturity breakdowns at three different levels which all are to be fixed in advance.

$$\lambda(m) = \begin{cases} a, \forall m \in [m_{min}; m_1[\\ b, \forall m \in [m_1; m_2[\\ c, \forall m \in [m_2; m_{max}[\end{cases}$$

B.3.2 Continuous variable roughness penalty function

Anderson and Sleath defined a continuous penalty function $\lambda(m)$ of m and three fixed parameters L, S and μ, which satisfy the following relationship:

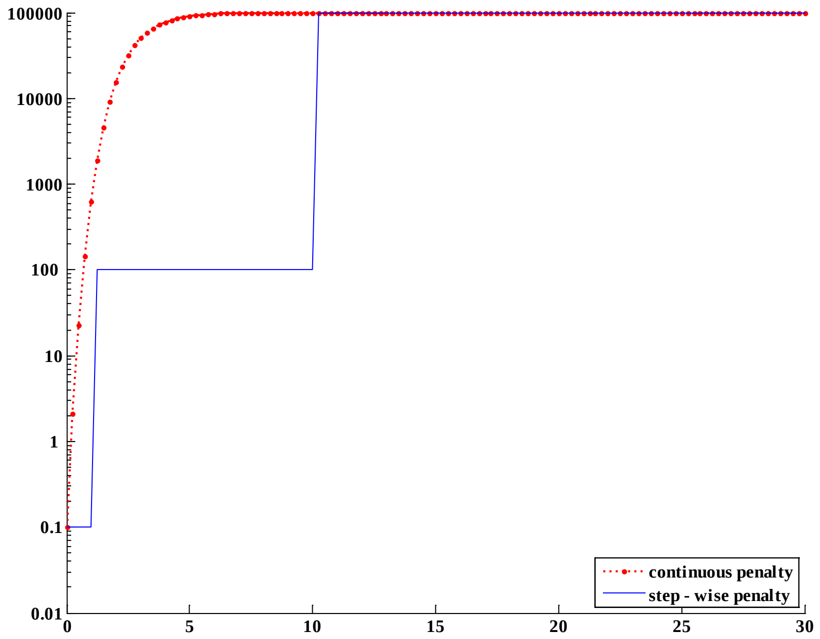
$$\log_{10} \lambda(m) = L - (L - S) \times \exp\left(-\frac{m}{\mu}\right)$$

Here is a plot of the two penalty functions, one for the continuous form (as used by the Bank of England) and one for the step-wise penalty (as used by the Federal Reserve Banks) using the following values:

Anderson and Sleath: L= 100,000; S= 0.1; μ= 1

Waggoner: a= 0.1; b= 100; c= 100,000; m1 = 1; m2 = 10

Chart B.1


100000
10000
1000
100
10
1
0.1
0.01
0
5
10
15
20
25
30
continuous penalty
step - wise penalty

C Appendix 3: Overview of the model properties for the parametric and spline-based models used for the empirical yield curve testing exercise

Four alternative derived methods are used for estimating the term structure, of which two are parametric-based models and the other two are spline-based models. An overview of the models and their properties, highlighting the main differences between the tests performed by the national central banks, is shown in the tables below.

Table C.1

Overview of the four methods and their properties for estimating the term structure used by the Banque de France

Property	Parametric models	
	Nelson & Siegel	**Svensson**
Nature of the forward rate curve	Forward rate is a single function defined at all maturities	
No of parameters	4	6
Objective function	Maximum likelihood (ML) minimisation price deviations weighted by duration $\min_{\alpha} \sum_{i=1}^{n} \frac{1}{D_i^2} \left(p_i(t,m_i) - \hat{p}_i(t,m_i,\alpha)\right)^2$ with duration $D_i^2 = \frac{\sum_s \frac{f_s \cdot t_s}{(1+r(t,m))^{t_s}}}{p(t,m)}$ cash flows fs ts=(time to cash flows)/365 Or minimisation yield deviations $\min_{\beta_t} \sum_{i=1}^{n_t} \left(r_{t,i}, -\hat{r}_{t,i}(\beta_t)\right)^2$	Maximum likelihood (ML) minimisation price deviations weighted by duration $\min_{\alpha} \sum_{i=1}^{n} \frac{1}{D_i^2} \left(p_i(t,m_i) - \hat{p}_i(t,m_i,\alpha)\right)^2$ with duration $D_i^2 = \frac{\sum_s \frac{f_s \cdot t_s}{(1+r(t,m))^{t_s}}}{p(t,m)}$ cash flows fs ts=(time to cash flows)/365 Or minimisation yield deviations $\min_{\beta_t} \sum_{i=1}^{n_t} \left(r_{t,i}, -\hat{r}_{t,i}(\beta_t)\right)^2$
Estimation algorithm	NLPCG: Conjugate Gradient optimisation method NLPDD: Double Dogleg optimisation method NLPNRA: Newton Raphson optimisation method NLPNRR: Newton Raphson Ridge optimisation method NLPTR: Trust Region optimisation method The above five SAS algorithms are used to estimate the parameters for price minimisation and the estimation result will be the best estimate of the five estimates. The result is a local optimum. These algorithms are standardised specific SAS algorithms: NLPCG: Non-Linear Conjugate Gradient optimisation method: during n iterations, the conjugate gradient algorithm computes a cycle of n conjugate search directions (Automatic Restart method of Powell (1977) and Beale (1972)), and find an approximate optimum of the function along the search direction (by quadratic interpolation and cubic extrapolation) NLPDD: Non-Linear Double Dogleg optimisation method: in each iteration, the algorithm computes the step, s(k), as a linear combination of the steepest descent or ascent search direction, s1(k), and a quasi-Newton search direction, s2(k), as follows: s(k)= α1 s1(k)+ α2.s2(k). The step s(k) must remain within a specified trust-region radius (refer to Fletcher, 1987). Hence, the NLPDD sub-routine uses the dual quasi-Newton update but does not perform a line search NLPNRA: Non-Linear Newton Raphson optimisation method: the NLPNRA algorithm uses a pure Newton step at each iteration when both the Hessian is positive definite and the Newton step successfully reduces the value of the objective function. Otherwise, it performs a combination of ridging and line search to compute successful steps. If the Hessian is not positive definite, a multiple of the identity matrix is added to the Hessian matrix to make it positive definite (refer to Eskow & Schnabel, 1991) NLPNRR: Non-Linear Newton Raphson Ridge optimisation method: the NLPNRR algorithm uses a pure Newton step when both the Hessian is positive definite and the Newton step successfully reduces the value of the objective function. Otherwise, a multiple of the identity matrix is added to the Hessian matrix NLPTR: Non-Linear Trust Region optimisation method: the NLPTR sub-routine is a trust-region method that uses the gradient and Hessian matrix. The trust-region method works by optimising a quadratic approximation of the non-linear objective function within a hyperelliptic trust region. This trust region has a radius, Δ, that constrains the	

Property	Parametric models	
	Nelson & Siegel	**Svensson**
	step size corresponding to the quality of the quadratic approximation. The method is implemented using Dennis, Gay, and Welsch (1981), Gay (1983), and Moré and Sorensen (1983) Or if yield minimisation is used: NLPNRR: Non-Linear Newton Raphson Ridge optimisation method is used	
		NLPFDD: Approximate Derivatives by Finite Differences: confidence level on β3 (the sub-routine returns the following values: f is a vector containing the values of the m functions comprising the objective function at the point x0. g is the m×n Jacobian matrix J, which contains the first-order derivatives of the functions with respect to the parameters, evaluated at x0. It is computed by finite difference approximations in a neighbourhood of x0. h is the n×n cross product of the Jacobian matrix, JTJ. It is computed by finite difference approximations in a neighbourhood of x0.)
Specified parameters	None	None
Constraints	Long-term asymptote β0 Parameters must be positive: β0 >0; β0+β1 = y1 >0; τ1 >0	Long-term asymptote β0 Parameters must be positive: β0 >0; β0+β1 = y1 >0; τ1 >0; τ2 >0
Starting values	Algorithm: scanning starting values set (β0, β1, β2, τ1) to find the global minimum (best estimation with the vector of parameters) by computing the multivariate density function fn.: $f_n(x) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{p}\frac{1}{h_j}K_j\left(\frac{x_{ij}-x_j}{h_j}\right)$ If the algorithm does not converge, the solution parameters from the day before are used.	Results of Nelson & Siegel parameters estimation are used as starting values for Svensson estimation. The initial starting value for β3 and τ2 are β3 = 1, τ2 = 0
Software used	SAS program developed by the Banque de France	SAS program developed by the Banque de France
Assessment criteria for same model test	Using the same starting values	

Table C.2

Overview of the two methods and their properties for estimating the term structure used by the Deutsche Bundesbank

Property	Parametric models	
	Nelson & Siegel	Svensson
Nature of the forward rate curve	Forward rate is a single function defined at all maturities	
No of parameters	4	6
Objective function	Minimisation of yield deviations $\min_{\beta_t} \sum_{i=1}^{n_t} \left(r_{t,i}, -\hat{r}_{t,i}(\beta_t)\right)^2$	Minimisation of yield deviations $\min_{\beta_t} \sum_{i=1}^{n_t} \left(r_{t,i}, -\hat{r}_{t,i}(\beta_t)\right)^2$
Estimation algorithm	Hanson/Krogh non-linear least squares with linear constraints based on quadratic-tensor local model (DQED) DQED internet: www.netlib.org/opt/dqed.f	
Specified parameters	None	None
Constraints	Long-term asymptote	Long-term asymptote
Starting values	β0=(yn+yn-1+yn-2) / 3 (mean yield of the three papers with the longest time-to-maturity) β1 = (y1 – β0); where y1 is the yield of the paper with the shortest time-to-maturity β2 =β3 = -1; where β3 is used as starting value for the Svensson model τ1 = τ2 = 1; where τ2 is used as starting value for the Svensson model If the algorithm does not converge the solution parameters from the day before are used	
Software used	Internal programming using Fortran-Routines	

Table C.3

Overview of the four methods and their properties for estimating the term structure used by the Bank of Greece

Property	Parametric models		Spline-based models
	Nelson & Siegel	Svensson	Waggoner step-wise linear penalty function
Nature of the forward rate curve	Spot rate is a single function defined at all maturities		Spot rate is piece-wise polynomial with segments joined at knot points
No of parameters	4	6	Number of knots + 2
Objective function	Minimise yield deviations Minimise $\sum_{i=1}^{N}[Y_i - \hat{Y}_i]^2$		Min $\sum_{i=1}^{N}[Y_i - \hat{Y}_i]^2 + \int_0^M \lambda(m)[f''(m)]^2 dm$ Where Lambda(m) is the step-wise penalty function
Estimation algorithm	Generalised reduced gradient method		Generalised reduced gradient method
Specified parameters	None	None	Number of knots is ~1/6 of total bonds spaced such that an equal number of bonds mature between knots For the step-wise penalty function the steps will be at residual maturities of one year and ten years
Constraints	Long-term asymptote	Long-term asymptote	None
Starting values	Solution parameters from the day before are used; manual estimation for first day		The average of the yield-to-maturity of all the bonds
Software used	Internal programming using C++,VB and Solver		Internal programming using C++,VB and Solver

Table C.4

Overview of the four methods and their properties for estimating the term structure used by the Bank of England

Property	Spline-based models	
	Waggoner step-wise linear penalty function	Waggoner continuous penalty function
Nature of the forward rate curve	Forward rate is piece-wise polynomial with segments joined at knot points	
No of parameters	Two more than the number of knot points (see below)	
Objective function	Minimise the fitted yields as the price divided by duration $\sum_{i=1}^{N}\left[\frac{P_i - \Pi_i(\beta)}{D_i}\right]^2 + \int_0^M \lambda(m)[f''(m)]^2 dm$	
	Where Lambda(m) is the step-wise function	Where Lambda(m) is a continuous function determined by an optimisation routine $\log\lambda(m) = L - (L - S)\exp\left(-\frac{m}{\mu}\right)$
Estimation algorithm	Essentially the first order approximation used in Fisher, Nychka and Zervos, except the objective function is price divided by duration and we have a continuous penalty function	
Specified parameters	Number of knots is 1/x of total bonds spaced such that an equal number of bonds mature between knots For the step-wise penalty function the steps will be at residual maturities of one year and ten years	The continuous penalty function is the one above, where the parameters L,S and μ are chosen to optimally fit the data, whilst not producing an "over-flexible" forward curve
Constraints	None	None
Starting values	The average of yield-to-maturity of all the bonds	
Software used	Programmed in MATLAB version 6.5	

D

Appendix 4: Yield curve statistics test results for the models for different time periods and fitting processes (yields or price minimisation)

Table D.1

In-sample goodness of fit statistics

	Yield curve statistics	period: 1/4/2003-31/3/2005			
	Model	Mean (MAE)	Standard deviation	Mean (RMSE)	Standard deviation
Yield	Nelson & Siegel (Yields) (Central bank of Greece)	0.04144	0.03698	0.05724	0.03953
	Nelson & Siegel (Yield) (Deutsche Bundesbank)	0.04415	0.01202	0.07104	0.02072
	Svensson (Yields) (Central bank of Greece)	0.03733	0.03657	0.04989	0.03196
	Svensson (Yield) (Deutsche Bundesbank)	0.04036	0.00680	0.06617	0.01789
	Nelson & Siegel/Svensson (Yield) (Banque de France)	0.01752	0.00179	0.02488	0.00191
	Waggoner with step-wise linear penalty function (Yield) (Central bank of Greece)	0.04479	0.04140	0.06124	0.03888
	Waggoner with continuous penalty function (Price) (Bank of England)	0.04768	0.00618	0.08971	0.01533
Price	Nelson & Siegel (Price) (Central bank of Greece)	0.20646	0.10531	0.41096	0.14403
	Nelson & Siegel (Price) (Deutsche Bundesbank)	0.22552	0.02055	0.44648	0.05370
	Svensson (Price) (Central bank of Greece)	0.17242	0.08881	0.32342	0.12026
	Svensson (Price) (Deutsche Bundesbank)	0.22207	0.02149	0.44552	0.05377
	Nelson & Siegel/Svensson (Price) (Banque de France)	0.01834	0.00266	0.02585	0.00481

Notes: The Bank of England has used “price weighted by duration” minimization. The Bank of Greece has minimised yield errors. From there, error statistics on both prices and yields were calculated.

Table D.2

Out-of-sample goodness of fit statistics

	Yield curve statistics	period: 1/4/2003-31/3/2005			
	Model	Mean (MAE)	Standard deviation	Mean (RMSE)	Standard deviation
Yield	Nelson & Siegel (Yields) (Central bank of Greece)	0.04051	0.03617	0.05562	0.03869
	Nelson & Siegel (Yield) (Deutsche Bundesbank)	0.04328	0.01568	0.06413	0.03807
	Svensson (Yields) (Central bank of Greece)	0.03706	0.03143	0.04915	0.03576
	Svensson (Yield) (Deutsche Bundesbank)	0.03958	0.01123	0.05894	0.03523
	Nelson & Siegel/Svensson (Yield) (Banque de France)	0.03398	0.01245	0.07214	0.03975
	Waggoner with step-wise linear penalty function (Yield) (Central bank of Greece)	0.03911	0.03340	0.05111	0.03150
	Waggoner with continuous penalty function (Bank of England)	0.04868	0.01465	0.08617	0.04431
Price	Nelson & Siegel (Price) (Central bank of Greece)	0.18131	0.09961	0.35893	0.13586
	Nelson & Siegel (Price) (Deutsche Bundesbank)	0.19358	0.05997	0.35945	0.13250
	Svensson (Price) (Central bank of Greece)	0.15573	0.07170	0.28823	0.09543
	Svensson (Price) (Deutsche Bundesbank)	0.19146	0.06115	0.36063	0.13482
	Nelson & Siegel/Svensson (Price) (Banque de France)	0.03549	0.01585	0.07440	0.04355

Notes: The Bank of England has used “price weighted by duration” minimization. The Bank of Greece has minimised yield errors. From there, error statistics on both prices and yields were calculated.

Table D.3

In-sample weighted mean absolute error (WMAE), RMSE and hit rates

Summary statistics			Period: 1/4/2003-31/3/2005								
			Maturities								
	Models	Statistics	All	0.25-1	]1-2	]2-3	]3-5	]5-7	]7-10	]10-20	]20-30
Yield	Nelson & Siegel (Central bank of Greece)	WMAE	0.04100	0.03600	0.03400	0.02300	0.03000	0.03700	0.05400	0.06800	0.06600
		RMSE	0.05700	0.05400	0.04400	0.03100	0.04100	0.04600	0.06800	0.08300	0.08800
		Hit rate	0.15017	0.21715	0.21894	0.17890	0.15930	0.07561	0.08181	0.04722	0.13022
	Nelson & Siegel (Deutsche Bundesbank)	WMAE	0.05330	0.05970	0.05300	0.03750	0.02870	0.03040	0.03940	0.05800	0.06540
		RMSE	0.07100	0.10280	0.06050	0.04280	0.03560	0.03890	0.05500	0.07610	0.08760
		Hit rate	0.11350	0.17240	0.09830	0.10520	0.11520	0.09760	0.09600	0.06530	0.08580
	Svensson (Central bank of Greece)	WMAE	0.03700	0.03200	0.03400	0.02300	0.02900	0.03700	0.05000	0.05900	0.04900
		RMSE	0.05000	0.04700	0.04300	0.03000	0.03900	0.04400	0.06100	0.07000	0.06700
		Hit rate	0.15591	0.23149	0.22408	0.19176	0.15174	0.06482	0.07031	0.05851	0.17500
	Svensson (Deutsche Bundesbank)	WMAE	0.04120	0.04500	0.03580	0.02570	0.02880	0.03770	0.04440	0.06150	0.06470
		RMSE	0.06620	0.09110	0.04650	0.03270	0.03640	0.04240	0.05370	0.07820	0.09160
		Hit rate	0.11940	0.18100	0.15860	0.19370	0.12300	0.03510	0.04160	0.05470	0.11770
	Nelson & Siegel/Svensson (Banque de France)	WMAE	0.01752	0.00205	0.01091	0.01829	0.01608	0.01412	0.02124	0.00889	0.00081
		RMSE	0.02488	0.00224	0.01253	0.02423	0.02533	0.01894	0.04086	0.02006	0.00102
		Hit rate	0.39690	0.92500	0.23810	0.28616	0.21716	0.16592	0.10282	0.15752	0.31416
	Waggoner with step-wise linear penalty function (Central Bank of Greece)	WMAE	0.04500	0.04200	0.03400	0.02300	0.03400	0.03900	0.05800	0.07300	0.07100
		RMSE	0.06100	0.06000	0.04300	0.03000	0.04500	0.04800	0.07400	0.08900	0.09200
		Hit rate	0.13065	0.18675	0.19267	0.15564	0.12903	0.06502	0.06872	0.03778	0.10678
	Waggoner with continuous penalty function (Bank of England)	WMAE	0.03326	0.07862	0.02621	0.03231	0.02339	0.01675	0.01502	0.02281	0.01718
		RMSE	0.08971	0.10676	0.04815	0.11298	0.09360	0.06272	0.06021	0.08906	0.07654
		Hit rate	0.29434	0.24929	0.40343	0.37590	0.42808	0.28288	0.32379	0.05725	0.04067
Price	Nelson & Siegel (Central bank of Greece)	WMAE	0.20646	0.01658	0.03395	0.04859	0.09484	0.18773	0.36543	0.63717	0.69409
		RMSE	0.41096	0.03863	0.05594	0.06358	0.14106	0.24379	0.48390	0.82907	0.92873
		Hit rate	0.15017	0.21715	0.21894	0.17890	0.15930	0.07561	0.08181	0.04722	0.13022
	Nelson & Siegel (Deutsche Bundesbank)	WMAE	0.07100	0.03380	0.05430	0.06450	0.09060	0.16340	0.27870	0.67660	1.06590
		RMSE	0.44650	0.05660	0.06900	0.07680	0.12290	0.20860	0.38060	0.84520	1.21920
		Hit rate	0.09450	0.12620	0.10440	0.10960	0.14370	0.06820	0.08030	0.01650	0.01520
	Svensson (Central bank of Greece)	WMAE	0.17242	0.01458	0.03280	0.04701	0.09218	0.18549	0.33889	0.53574	0.50451
		RMSE	0.32342	0.03138	0.05479	0.06116	0.12933	0.23676	0.43223	0.65979	0.69549
		Hit rate	0.15591	0.23149	0.22408	0.19176	0.15174	0.06482	0.07031	0.05851	0.17500
	Svensson (Deutsche Bundesbank)	WMAE	0.06520	0.02790	0.04240	0.05950	0.09080	0.16990	0.27480	0.67080	1.06810
		RMSE	0.44550	0.05110	0.06140	0.07130	0.12090	0.21060	0.37930	0.84250	1.21950
		Hit rate	0.10550	0.13800	0.17990	0.11070	0.12750	0.05750	0.09080	0.01660	0.01520
	Nelson & Siegel/Svensson (Banque de France)	WMAE	0.01833	0.00482	0.01094	0.01396	0.01628	0.00907	0.01234	0.00674	0.00081
		RMSE	0.02584	0.00543	0.01319	0.02196	0.03653	0.01563	0.02491	0.01879	0.00102
		Hit rate	0.31510	0.95276	0.72237	0.43135	0.38669	0.33064	0.23806	0.24810	0.31416

Notes: The Bank of England has used “price weighted by duration” minimization. The Bank of Greece has minimised yield errors. From there, error statistics on both prices and yields were calculated.

Table D.4

Out-of-sample weighted mean absolute error (WMAE), RMSE and hit rates

	Summary statistics		Period: 1/4/2003-31/3/2005								
			Maturities								
	Models	Statistics	All	0.25-1	]1-2	]2-3	]3-5	]5-7	]7-10	]10-20	]20-30
Yield	Nelson & Siegel (Central bank of Greece)	WMAE	0.04100	0.03700	0.03600	0.02500	0.03000	0.03700	0.05400	0.06600	0.06100
		RMSE	0.05600	0.05500	0.04700	0.03300	0.04000	0.04500	0.06800	0.08000	0.08400
		Hit rate	0.14714	0.20543	0.21030	0.18465	0.15305	0.07235	0.08512	0.04114	0.11867
	Nelson & Siegel (Deutsche Bundesbank)	WMAE	0.05390	0.06070	0.05340	0.03670	0.02890	0.02950	0.03870	0.05540	0.06000
		RMSE	0.06410	0.07690	0.05750	0.03990	0.03390	0.03390	0.04840	0.06190	0.06840
		Hit rate	0.11460	0.18370	0.09250	0.10460	0.10840	0.10220	0.08830	0.05100	0.08480
	Svensson (Central bank of Greece)	WMAE	0.03700	0.03300	0.03500	0.02500	0.02900	0.03700	0.05000	0.05900	0.04400
		RMSE	0.04900	0.04800	0.04500	0.03300	0.03800	0.04400	0.06100	0.06900	0.06300
		Hit rate	0.14920	0.21177	0.21289	0.19757	0.14927	0.06088	0.07438	0.04517	0.15156
	Svensson (Deutsche Bundesbank)	WMAE	0.04230	0.04740	0.03630	0.02600	0.02880	0.03760	0.04350	0.05860	0.05930
		RMSE	0.05890	0.06580	0.04170	0.03040	0.03430	0.04050	0.04960	0.06400	0.06850
		Hit rate	0.12070	0.16680	0.15300	0.21100	0.12490	0.03360	0.03850	0.05130	0.14060
	Nelson & Siegel/Svensson (Banque de France)	WMAE	0.03398	0.02548	0.00586	0.00366	0.01126	0.01175	0.01270	0.01671	0.00354
		RMSE	0.07214	0.02914	0.00796	0.00466	0.01283	0.01351	0.01557	0.03598	0.00354
		Hit rate	0.34811	0.57778	0.66667	0.50000	0.00000	0.15897	0.20313	0.16705	0.00000
	Waggoner with step-wise linear penalty function (Central Bank of Greece)	WMAE	0.03900	0.03700	0.03500	0.02500	0.03200	0.03800	0.05200	0.06100	0.04400
		RMSE	0.05100	0.05100	0.04500	0.03200	0.04200	0.04600	0.06500	0.07100	0.06000
		Hit rate	0.13243	0.18078	0.18506	0.15695	0.12550	0.06294	0.07491	0.03703	0.10206
	Waggoner with continuous penalty function (Bank of England)	WMAE	0.03569	0.08038	0.02694	0.03110	0.02193	0.01238	0.01928	0.01662	0.00251
		RMSE	0.08617	0.08070	0.04137	0.07003	0.06577	0.03677	0.06794	0.06427	0.01154
		Hit rate	0.30026	0.22296	0.36013	0.32786	0.40803	0.30528	0.20932	0.02861	0.00697
Price	Nelson & Siegel (Central bank of Greece)	WMAE	0.18131	0.01719	0.03410	0.05115	0.09400	0.18690	0.36298	0.58368	0.65673
		RMSE	0.35893	0.03839	0.05402	0.06760	0.13837	0.24005	0.48536	0.74326	0.88210
		Hit rate	0.14714	0.20543	0.21030	0.18465	0.15305	0.07235	0.08512	0.04114	0.11867
	Nelson & Siegel (Deutsche Bundesbank)	WMAE	0.06980	0.03460	0.05600	0.06470	0.09280	0.16230	0.27680	0.63760	0.97920
		RMSE	0.35940	0.04520	0.06400	0.07250	0.11600	0.18560	0.34550	0.68490	1.05270
		Hit rate	0.09460	0.12360	0.09660	0.11930	0.13730	0.06310	0.06720	0.01360	0.00880
	Svensson (Central bank of Greece)	WMAE	0.15573	0.01477	0.03298	0.04967	0.09206	0.18622	0.33675	0.49883	0.46692
		RMSE	0.28823	0.02741	0.05322	0.06594	0.12678	0.23613	0.43324	0.59979	0.64523
		Hit rate	0.14920	0.21177	0.21289	0.19757	0.14927	0.06088	0.07438	0.04517	0.15156
	Svensson (Deutsche Bundesbank)	WMAE	0.06460	0.02900	0.04480	0.06050	0.09240	0.16980	0.27430	0.63780	0.98830
		RMSE	0.36060	0.03960	0.05420	0.06770	0.11410	0.19120	0.34470	0.68570	1.06180
		Hit rate	0.10430	0.13460	0.17020	0.11570	0.12730	0.05220	0.07740	0.01560	0.00830
	Nelson & Siegel/Svensson (Banque de France)	WMAE	0.03549	0.01298	0.01447	0.00906	0.00681	0.01211	0.00209	0.01444	0.00014
		RMSE	0.07440	0.01304	0.01489	0.00938	0.00809	0.01475	0.00234	0.02626	0.00014
		Hit rate	0.28885	0.96667	0.92308	0.89951	0.85242	0.72777	0.80319	0.29469	0.89712

Notes: The Bank of England has used “price weighted by duration” minimization. The Bank of Greece has minimised yield errors. From there, error statistics on both prices and yields were calculated.

E

Appendix 5: Visual presentation of the daily estimations of yield curves covering the two-year dataset period

Video demo of the daily Svensson model yield curve estimations (Bank of Greece)

Video demo of the daily Svensson model yield curve estimations (Deutsche Bundesbank)

Video demo of the daily Svensson and Nelson & Siegel switching model yield curve estimations (Banque de France)

Video demo of the daily Waggoner model with continuous penalty model yield curve estimations (Bank of England)

Link: http://www.ecb.europa.eu/pub/pdf/annex/ecb.sps27_annex5.en.zip

Appendix 6: Model properties for calculating euro area yield curves

Table F.1

Overview of the ECB methodology for implementing the Svensson model

Parametric model	Svensson
Type of curve	Three types of nominal curves (the spot yield curve, the instantaneous forward curve and the par yield curve)
No of parameters	Six
Starting values	Use of previous day's parameter values for today's starting values
Constraints	Long-term asymptote β0. Parameters must be positive
Minimisation	Minimisation of yield deviations
Yield calculations	ISMA formula 6.4 [1]
Day-count conventions	Depending on type of bonds and issuer location. Following market practice such as Actual/360, Actual/Actual or 30E/360 ISMA
Calculation start	Calculations based on settlement days taking into account first subsequent business day according to individual bond market practice
Sector coverage	Euro area central government (ESA S.1311). No supranational issuers
Denomination	Euro
Type of coupon	Zero- and fixed-coupon bonds including strips. No adjustments for tax or coupon effects. Exclusion of bonds with special features (e.g. inflation-linked)
Maturity spectrum	Three months to thirty years
Issuer ratings	AAA issuer country rating provided by Fitch, and other ratings
Price selection	Latest executed daily price/mid-quote at close of markets
Liquidity criteria	Minimum trading volume of EUR 1 million and above on any given day. Testing results indicate daily trading volumes of EUR 20-100 million. Bid-ask spread of a maximum of three basis points
Data source	EuroMTS and Fitch
Outlier removal	Yes, two-layer outlier removals applied. Defined as bonds with a yield more than twice the standard deviation of the maturity bracket average
Daily production	Business daily (following the TARGET calendar)

1) Fully paid fixed-coupon bond with an assumed single redemption date, using the settlement date; ISMA formulas for yield and other calculations (the International Securities Market Association, 1992).

List of abbreviations

bps	basis points
BdF	Banque de France
BIS	Bank for International Settlements
BoE	Bank of England
BoG	Bank of Greece
CET	Central European Time
ECB	European Central Bank
EONIA	euro overnight index average
ESA 2010	European System of National and Regional Accounts
ESCB	European System of Central Banks
Eurepo	interest rate for secured money market transactions in the euro area
EURIBOR	euro interbank offered rate
FNZ	Fisher, Nychka and Zervos
MAE	mean absolute error
N&S	Nelson & Siegel (model)
OTC	over-the-counter
RMSE	root mean squared error
S.11	non-financial corporations (ESA sector)
S.121+S.122+S.123	monetary financial institutions (the ECB and the national central banks of the euro area (S.121), deposit-taking corporations (S.122) and money market funds (S.123)) (ESA sector)
S.125	other financial intermediaries (ESA sector)
S.1311	central government (excluding social security funds) (ESA sector)
S.1312	state government (ESA sector)
S.1313	local government (ESA sector)
S.1314	social security funds (ESA sector)
VRP	variable roughness penalty (function)
WMAE	weighted mean absolute error

References

Anderson et al. (1996), "Estimating and Interpreting the Yield Curve", p. 25.

Anderson and Sleath (2001), "New estimates of the UK real and nominal yield curves", Working Paper, Bank of England.

Berk, J.M. and Van Bergeijk, P.A.G. (2000), "Is the Yield Curve a useful information variable for the Eurosystem?", *Working Paper Series*, No 11, ECB.

Bliss, R.R. (1996), "Testing Term Structure Estimation Methods", Working Paper 96-12a, Federal Reserve Bank of Atlanta.

Bolder, D. and Stréliski, D. (1999), "Yield curve modelling at the Bank of Canada", Technical Report 84, Bank of Canada.

Bolder, D. and Gusba, S. (2002), "Exponentials, polynomials and Fourier series: more yield curve modelling at the Bank of Canada", Working Paper 2002-99, Bank of Canada.

Brockwell, P.J. and Davis, R.A. (2016), *Introduction to Time Series and Forecasting*, Springer International Publishing, Switzerland.

Brousseau, V. (2002), "The functional form of yield curves", *Working Paper Series*, No 148, ECB.

Cassola, N. and Barros, L.J. (2001), "Two-factor model of the German term structure of interest rates", *Working Paper Series*, No 46, ECB.

De Boor, C. (1978), "A practical guide to splines".

Dunne, P., Moore, M.J. and Portes, M. (2014), "Price discovery in the European bond market", London Business School, December 2004.

ECB (2008), "The new euro area yield curves", *Monthly Bulletin*, February.

Fisher, M., Nychka, D. and Zervos, D. (1994), "Fitting the Term Structure of Interest rates with Smoothing Splines", Working Paper, Federal Reserve Bank of Atlanta.

Garvin, C.F. (1996), *Financial Toolbox: For Use with MATLAB*, Natick, Mass.: MathWorks.

Krippner, L. (2015), *Zero Lower Bound Term Structure Modeling: A Practitioner's Guide*, Applied Quantitative Finance, Palgrave Macmillan.

McCulloch, J. (1971), "Measuring the Term Structure of Interest Rates", *Journal of Business 44*, 19–31.

Nelson, C.R. and Siegel, A.F. (1987), "Parsimonious Modeling of Yield Curves", *Journal of Business 60*, 473–489.

Neumayer, E. and Plümper, T. (2017), *Robustness Tests for Quantitative Research*, Cambridge University Press.

Nymand-Andersen, P., Golding, M. and Moreno, A.-J. (2004), "Measuring prices from financial instruments: Yield curve modelling", *Bank for International Settlements*, Irving Fisher Committee, Bulletin No 19.

Ruppert, D. and Caroll, F.J. (2000), "Spatially-adaptive penalties for spline fitting".

Schich, S. (1997), "Estimating the German term structure", Discussion paper 4/97, Economic Research Group of the Deutsche Bundesbank.

Svensson, L.E. (1994), "Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994", Discussion Paper 1051, Centre for Economic Policy Research.

Svensson, L.E. and Söderlin, P. (1997), "New Techniques to Extract Market Expectations from Financial Instruments", Discussion Paper 1556, Centre for Economic Policy Research.

Vasicek, O.A. and Fong, H.G. (1982), "Term Structure Modelling Using Exponential Splines", *Journal of Finance 37*, 339–348.

Waggoner, D.F. (1997), "Spline Methods for Extracting Interest Rate Curves from Coupon Bond Prices".

Wegman, E.J. and Wright, I.W. (1983), "Splines in Statistics".

Acknowledgements

I would like to thank Antonio Moreno, Patrick Markowicz and Vasileios Georgakopoulos (ECB), Polychronis Manousopoulos (Bank of Greece), Giovanni Guazzarotti (Banca d'Italia), Fahd Rachidy and Maryam Housni Fellah (Banque de France), Matthew Hurd, Andrew Meldrum and James Mason (Bank of England), and Jörg Meier and Jelena Stapf (Deutsche Bundesbank), for their substantial assistance in the empirical exercise. Without their contribution, this paper would not have been possible. I would also like to thank the ECB's Statistics Paper Series Editorial Board for its valuable input to the paper.

Per Nymand-Andersen

European Central Bank, Frankfurt am Main, Germany; email: per.nymand@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

and from RePEc: Research Papers in Economics. Information on all of the papers published in the ECB Statistics Paper Series can be found on the ECB's website.

ISSN	2314-9248 (pdf)	DOI	10.2866/892636 (pdf)
ISBN	978-92-899-3356-8 (pdf)	EU catalogue No	QB-BF-18-001-EN-N (pdf)