

Econometric Tools

Ken Nyholm

Outline

- Nelson-Siegel yield curves
- Vector AutoRegression (VAR) models
- Regime-switching

Nelson-Siegel yield curves

- Parametric yield curve models
- Nelson-Siegel vs affine yield curve models
- Non-parametric yield curve models
- Short rate models
- Absence of arbitrage

Nelson-Siegel yield curves

- The Nelson-Siegel model in yield form:

Following the intuitive parametrisation suggested by Diebold and LI (2006) the functional form for the Nelson-Siegel model is:

$$Y_t(\tau) = H \cdot \beta_t + e_t, \quad (4.1)$$

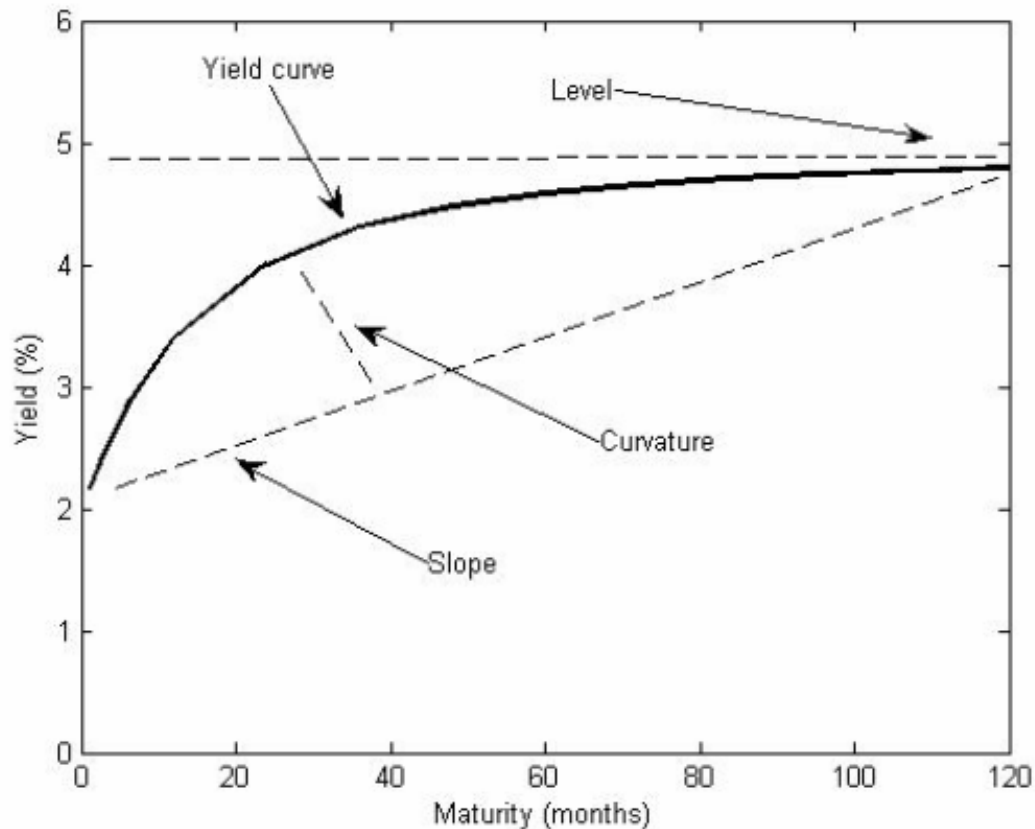
where $Y_t(\tau)$ is the vector of yields observed at time t for the T maturities τ . The matrix of factor sensitivities is explicitly given by:

$$H = \begin{bmatrix} 1 & \frac{1 - \exp(-\lambda\tau_1)}{\lambda\tau_1} & \frac{1 - \exp(-\lambda\tau_1)}{\lambda\tau_1} - \exp(-\lambda\tau_1) \\ 1 & \frac{1 - \exp(-\lambda\tau_2)}{\lambda\tau_2} & \frac{1 - \exp(-\lambda\tau_2)}{\lambda\tau_2} - \exp(-\lambda\tau_2) \\ \vdots & \vdots & \vdots \\ 1 & \frac{1 - \exp(-\lambda\tau_T)}{\lambda\tau_T} & \frac{1 - \exp(-\lambda\tau_T)}{\lambda\tau_T} - \exp(-\lambda\tau_T) \end{bmatrix}. \quad (4.2)$$

The yield curve factors are collected in $\beta_t = \{\text{level}_t, \text{slope}_t, \text{curvature}_t\}^T$. λ determines the speed of time-decay for the slope and curvature sensitivities, i.e.

Nelson-Siegel yield curves

- The yield curve factors: $\beta_t = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}_t = \begin{bmatrix} \text{yield curve level} \\ \text{yield curve slope} \\ \text{yield curve curvature} \end{bmatrix}_t$.



Nelson-Siegel yield curves

- The yield curve sensitivities

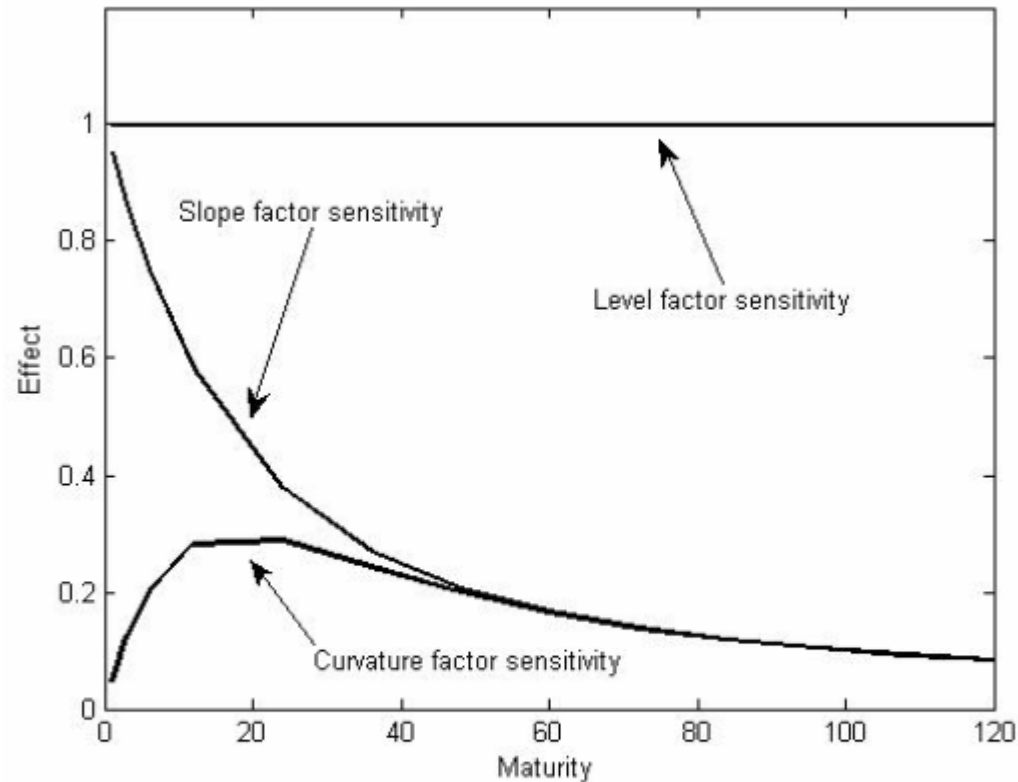


Figure 4.2: Nelson-Siegel factor sensitivities ($\lambda = 0.08$)

Nelson-Siegel yield curves

- Set in state space form:

The observation equation

The vector of yields Y at time t for different maturities $\tau = \{\tau_1, \tau_2, \dots, \tau_n\}$ can be expressed as a function of yield curve factors and yield curve factor sensitivities.

$$Y_t(\tau) = H\beta_t + \varepsilon_t, \quad (4.8)$$

where H collects the Nelson-Siegel factor sensitivities:

Nelson-Siegel yield curves

The state equation

The state equation describes how the unobserved factors evolve over time. Here it is assumed that a VAR(1) with regime switching in the means is appropriate. Hence:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}_t = \begin{bmatrix} m_1^I \\ m_2^J \\ m_3^K \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}_{t-1} + v_t, \quad (4.11)$$

where $v \sim N(0, Q)$. The superscript on the mean parameters indicate regime affiliation. In our example of the model, however, only the slope is allowed to exhibit regime switching behaviour i.e. $I = \{I\}$, $J = \{a, b, c\}$, $K = \{K\}$. The

Nelson-Siegel yield curves

- Exercise: Estimate the parameters of the Nelson-Siegel yield curve model on the following data:

```
Y = [2.17 2.47 2.85 3.39 4.01 4.31 ...  
     4.48 4.58 4.65 4.70 4.74 4.77 4.79]';  
tau = [1 3 6 12 24 36 48 60 72 84 96 108 120]';
```

VAR models

The future evolution of most financial and economical time series depend on their own past evolution and cross correlation with past observations of other variables. Vector Autoregressive models exploit such dependency structures by characterising the observation at time t of a vector X as a function of past observations $\{X_{t-1}, X_{t-2}, \dots, X_{t-j}\}$. The process for the mean of the vector series is thus:

$$X_t = c + \sum_{j=1}^p X_{t-j} * A_j + e_t. \quad (4.3)$$

This is called a $VAR(p)$ model where p refers to the number of lags included in the model, and A_j collects the autoregressive parameters at lag j . In the formulation above, X_t is a matrix of dimension $(nObs - p) \times nVars$, where $nVars$ refers to the number of variables contained in X . The optimal lag-length can be determined by the use of summary statistics that are based on a specific weighting of the fit of the model against the number of parameters that are necessary to obtain this fit. Hence, these statistics gives a positive weight to the degree of fit obtained and includes a penalty for the number of variables needed to obtain the fit. In this way a parsimonious model can be found, i.e. the one that uses the fewest variables to get the best fit, in a relative sense. Parsimonious models have proven particularly useful for forecasting: it is seldom the case that

VAR models

Two of such statistics are the Akaike's and Schwarz Information Criterion AIC and BIC, respectively. These can be calculated in the following way²:

$$AIC = \log(\hat{\sigma}^2) + 2 * \frac{k}{nObs} \quad (4.4)$$

$$BIC = \log(\hat{\sigma}^2) + k * \frac{\log(T)}{nObs} \quad (4.5)$$

It can be seen from (4.4) and (4.5) that the fit of the model is assessed by the log of the sum of squared residuals, and then each criterion adds a penalty for the number of parameters used to obtain this fit: k represents the number of estimated parameters and $nObs$ the number of observations.

VAR models

Related to (4.3) it is worth noting that any VAR(p) model can be written as a VAR(1) model: this is also called to write the VAR(p) model in companion form. Assume that we look at a VAR(3) model i.e.

$$X_t = c + X_{t-1} * A_1 + X_{t-2} * A_2 + X_{t-3} * A_3 + e_t.$$

By defining

$$\tilde{X}_t = \begin{bmatrix} X_t \\ X_{t-1} \\ X_{t-2} \end{bmatrix}$$

and

$$\tilde{A} = \begin{bmatrix} A_1 & A_2 & A_3 \\ I & 0 & 0 \\ 0 & I & 0 \end{bmatrix},$$

where I is the identity matrix of appropriate dimension. Then the VAR(3) can be written as a VAR(1):

$$\tilde{X}_t = c + \tilde{A} * \tilde{X}_{t-1} + e_t. \quad (4.6)$$

VAR models

The integration orders typically seen in finance and economics are of orders of 0 and 1, written as $I(0)$ and $I(1)$, respectively. One rarely encounters time series of order 2 or higher. $I(0)$ refers to a stationary process and $I(1)$ to a process that is integrated of order 1. Hence, if Y_t is $I(1)$, then the process $\Delta Y_t = Y_t - Y_{t-1}$ will be $I(0)$. A standard example of an $I(1)$ process is:

$$x_t = k + a * x_{t-1} + e_t,$$

with $a = 1$; this is the well-known random walk model. Stationarity in the univariate case is defined by the value of a . When $-1 < a < 1$, the process above is $I(0)$. For a process involving more lags than one it is a bit more complicated to determine whether a process is stationary or not. The requirements are listed here, for detail see Hamilton (1994): For an $AR(p)$ model it is required that the roots of the following polynomial all lie outside the unit circle:

$$1 - a_1 * z - a_2 * z^2 - \dots - a_p z^p = 0.$$

Similarly, for the $VAR(p)$ model it is required that the roots of the following polynomial all lie outside the unit circle:

$$\det [I - A_1 * z - A_2 * z^2 - \dots - A_p * z^p] = 0,$$

where \det indicates the determinant. Given that the process at hand is station-

VAR models

where \det indicates the determinant. Given that the process at hand is stationary its mean is calculated by:

$$E[X] = \frac{c}{I - \sum_{j=1}^p A_j},$$

for the $VAR(p)$ model and by:

$$E[x] = \frac{c}{1 - \sum_{j=1}^p a_j}$$

for the $AR(p)$ model.

VAR models

It can be shown that the VAR(p) model can be estimated efficiently by the use of consecutive OLS regressions. Alternatively one can choose to maximise the value of the multivariate log-likelihood function assuming that errors are normally distributed. This amounts to calculating the approximative log likelihood of the multivariate normal distribution function, as given by:

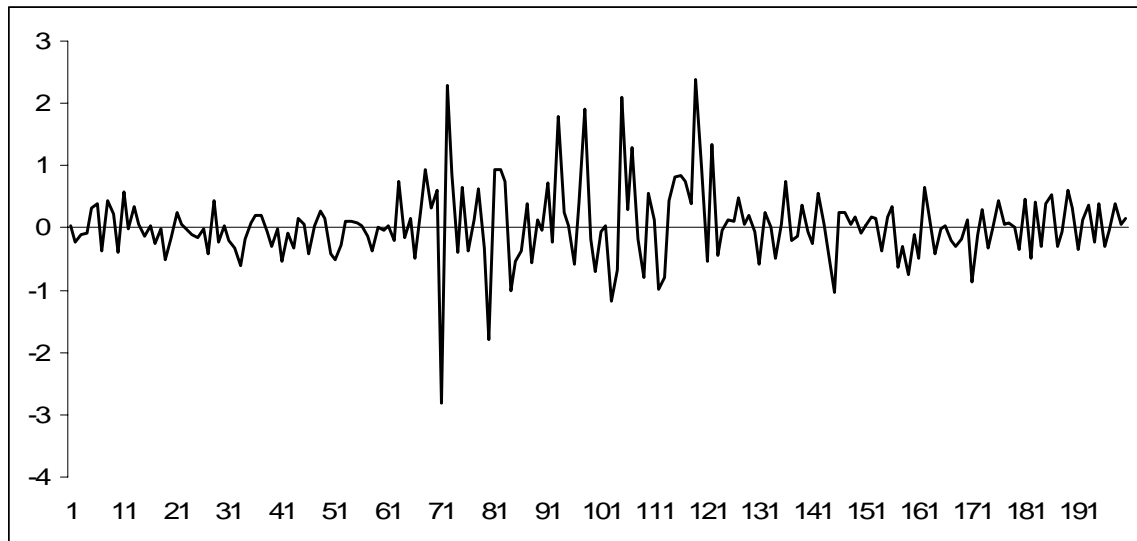
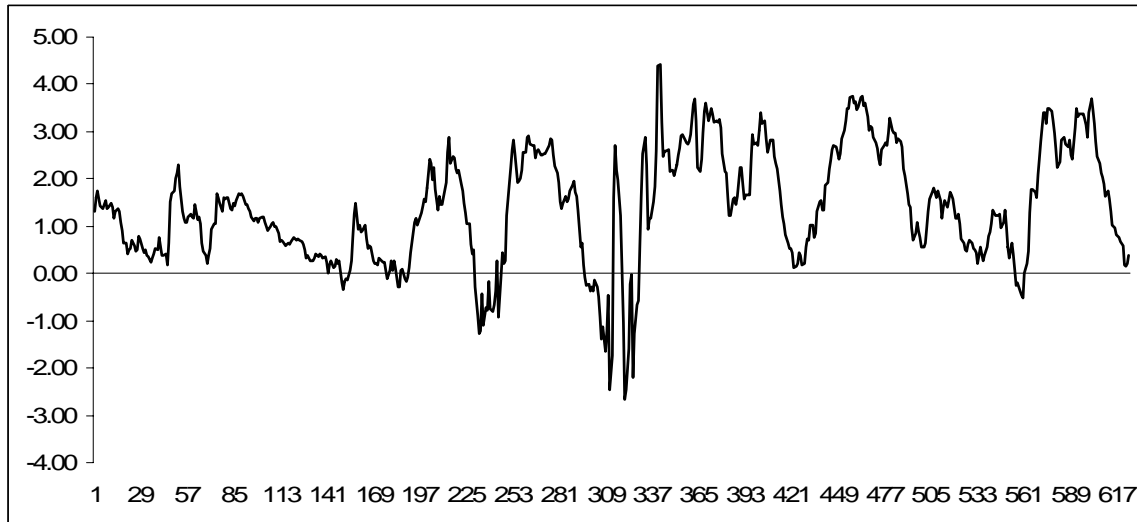
$$\log L = \frac{nObs - p}{2} * \log [\det (\Omega^{-1})] - \frac{1}{2} * \sum_{j=p+1}^{nObs} e_t * \Omega^{-1} * e_t^T.$$

In terms of speed however, since closed-form expressions exist for the solution to the OLS regression, this is typically faster than it is to optimise the log likelihood function.

VAR models

- An example of estimating a VAR using the log likelihood function is given in *var_p.m* which calls *var_p_likeli.m*

Regime-switching



Regime-switching

The intuition behind regime-switching models is perhaps best illustrated via a standard linear dummy-variable regression of the form:

$$x_t = c_1d_{1,t} + c_2d_{2,t} + c_3d_{3,t} + c_4d_{4,t} + e_t.$$

Here x is the data thought to contain two or more states, s indicates the number of states, the c 's are constants that define the mean in each state. The d variables represent dummy variables defining when a given state is effective. It is further assumed that the error term is normally distributed: $e \sim N(0, \sigma^2)$. In a linear regression model usually the dummy variables are fixed exogenously, however, in a regime-switching model the d 's as well as the c 's are inferred from data.⁴

Regime-switching

In order to facilitate estimation of these variables, where the d 's are going to be treated as unobservable, a prediction-updating algorithm is applied to construct the likelihood function. To set up this algorithm a number of inputs need to be defined. These are: starting values for the parameters to be estimated, which in our example is: $\theta = \{c_1, c_2, c_3, c_4, \sigma^2, P\}$, and a transition probability matrix P , that gives the conditional probability of migrating within and between states. For example, in the three state case P would be:

$$P = \begin{bmatrix} p_{11} & p_{21} & 1 - p_{33} - p_{32} \\ p_{12} & p_{22} & p_{32} \\ 1 - p_{11} - p_{12} & 1 - p_{22} - p_{21} & p_{33} \end{bmatrix}.$$

The entries in this matrix give the probabilities for one state being followed by another state, for example, the third row, second column entry (p_{32}) gives the probability that state 2 is followed by state 3. Consequently, columns must sum

Regime-switching

chain with P as transition matrix. These state probabilities are collected in $\pi_{t|t}$, which is of dimension $[s \times 1]$ and gives the probability that the process is in a given state at time t . For example, in the three state case, it could be that:

$$\pi_{t|t} = \begin{bmatrix} 0.20 \\ 0.70 \\ 0.10 \end{bmatrix},$$

with each entry in π gives the probability that the datapoint observed at t is generated by that particular state. The slightly peculiar subscript comes from the prediction-updating algorithm described below. The matrix of values of the density for each observation, within each state, is collected in $D = f(x_t | s_t = j; \theta)$ for all observations t and states j . Here $f(\cdot)$ is taken to be the normal distribution, hence, again using the three state case as an example,

$$D_t = \begin{bmatrix} \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ \frac{-(x_t - c_1)^2}{2\sigma^2} \right\} \\ \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ \frac{-(x_t - c_2)^2}{2\sigma^2} \right\} \\ \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ \frac{-(x_t - c_3)^2}{2\sigma^2} \right\} \end{bmatrix}.$$

Regime-switching

The Hamilton filter then runs according to [see Hamilton (1994, Chapter 22)]:

$$\begin{aligned}\pi_{t|t} &= \frac{\pi_{t|t-1} \odot D_t}{\sum_{j=1}^s \pi_{t|t-1} \odot D_t}, \\ \pi_{t+1|t} &= P\pi_{t|t}.\end{aligned}$$

Iterating over all $t = 1, 2, \dots, T$ for $\pi_{t|t}$ and $\pi_{t+1|t}$ is thus possible; \odot indicates element-by-element multiplication. The intuition of the expression for $\pi_{t|t}$ is that the numerator is the product of the probability forecast based only on the mechanical forwarding mechanism (i.e. information up until $t - 1$) given by the expression for $\pi_{t+1|t}$ and the density of the observation at t for each state. Hence, the numerator is a $[s \times 1]$ vector of the probability weighted fits of the t 'th observation to each of the hypothesized state specifications. The denominator, which is the sum over the number of states of the numerator, can be seen as to normalize the numerator so that each value in the vector $temp = \pi_{t|t-1} \odot D_t$ falls on the interval $\{0, \dots, 1\}$ and thus can be interpreted as probabilities. It so

Regime-switching

falls on the interval $\{0, \dots, 1\}$ and thus can be interpreted as probabilities. It so happens, that the denominator also is identical to the likelihood function value conditional on θ , and the estimation can thus be completed by optimising:

$$\log L(\theta) = \sum_{t=1}^T \log \left\{ \sum_{j=1}^s \pi_{t|t-1} \odot D_t \right\}.$$

In practise it is not always easy to find the number of regime-switches that are contained in data. Naturally, one needs to plot data before modelling it, and this can be used as a first indication for the choice on the number of states to be included. Another, equally valid option, is to be guided by economic theory,

Regime-switching

- To estimate a regime-switching model the following is needed:
 - 1) A main function that sets up the problem and calls the Matlab(tm) optimisation module. This function is called *regime.m*
 - 2) The likelihood function is calculated by *regime_likeli.m*. This function calls two other functions: *regime_normaldensity* and *regime_pmat*.
 - 2a) *regime_normaldensity* simply implements the density function of the normal distribution.
 - 2b) *regime_pmat* sets up the transition matrix shown in () so that it matches the desired number of states to be estimated.
- This is implemented in *regime.m*,
regime_likeli.m, *regime_normaldensity.m*,
regime_pmat.m