A theoretical foundation for the Nelson and Siegel class of yield curve models

Leo Krippner*

June 5, 2009

Abstract

This article establishes that yield curve models within the popular Nelson and Siegel (1987) class are effectively reduced-form representations of the generic Gaussian affine term structure model outlined in Dai and Singleton (2002). That provides a fundamental theoretical foundation for Nelson and Siegel (1987) models, and a compelling case for applying them as standard standard tools for yield curve analysis in economics and finance.

JEL: E43, G12

Keywords: yield curve; term structure of interest rates; Nelson and Siegel model; affine term structure models.

1 Introduction

This article establishes that yield curve models within the Nelson and Siegel (1987, hereafter NS) class are effectively reduced-form representations of the generic Gaussian affine term structure model (hereafter GATSM) outlined in Dai and Singleton (2002). That exposition initially provides a long-overdue fundamental theoretical foundation for NS models. As background, NS models have become widely used in economics and finance based on their practical benefits and empirical successes,[1] but the Level, Slope, and Curvature factor loadings at the core of all NS models have their origin in the somewhat arbitrary and atheoretical field of yield curve fitting.[2] An explicit correspondence with the GATSM framework provides users with an assurance that NS models conform to a well-accepted set of principles and assumptions for modelling the yield curve.

The corollary that follows from the exposition is a compelling case for applying NS models instead of GATSM models as standard tools for routine yield curve analysis in economics and finance. That is, unless any of the unique state variables and/or parameters of the GATSM are explicitly of interest, the user can bypass the relative complexity of specifying, identifying, estimating, and interpreting the GATSM by simply applying an appropriate NS model. The NS model will provide the same practical results in terms of summarising the shape of the yield curve and its evolution over time, and the estimated NS coefficients and parameters will still provide an economic and financial meaning in the context of the application.

The exposition begins in section 2 by specifying the generic GATSM outlined in Dai and Singleton (2002) within the context of this article. The forward rate curve associated with the GATSM is then derived, and section 3 explicitly shows how that forward rate curve may be re-expressed in the original NS representation. Specifically, the Level factor loading and its associated coefficient are shown to correspond to the highly-persistent (i.e very slowly mean-reverting) components of the

*Reserve Bank of New Zealand (email: leo.krippner@rbnz.govt.nz).

[1] Bank for International Settlements (2005) provides an overview of central bank use, and recent examples include Gürkaynak, Sack and Wright (2008), Diebold, Li and Yue (2008), Christensen, Diebold and Rudebusch (2008), and Yu and Salyards (2009).

[2] The forward rate curve of the original NS model was simply proposed as the general solution to a second-order differential equation with equal roots, and then integrated to obtain the NS interest rate curve.

GATSM, and the Slope and Curvature factor loadings with their associated coefficients are shown to correspond to the non-persistent (i.e mean-reverting) components of the GATSM. In light of this example, section 4 discusses how models within the NS class can be classified as various representations of a generic GATSM. This provides some guidance on selecting the appropriate NS model and for interpreting its output.

2 The generic Gaussian affine term structure model

The generic GATSM specified in this section mainly follows Appendix A of Dai and Singleton (2002),[3] but it is worth highlighting two points of context for this article. First, the state variables themselves are completely generic, so they can be taken to represent points on the term structure as suggested by Duffie and Kan (1996), and/or economic and financial factors within the underlying economy.[4] Second, to make the exposition more transparent from the perspective of the original NS model, this article generally works with the forward rate curve associated with the generic GATSM rather than the bond prices and/or interest rate curve typically used in affine term structure models.

Define the instantaneous short rate at time t as $r(t) = \rho_0 + \rho_1' X(t)$ where ρ_0 is a constant, $X(t)$ is a $N \times 1$ vector of state variables, and ρ_1 is a constant $N \times 1$ vector. Under the physical P measure, the state variables follow the process $dX(t) = K_P [\theta_P - X(t)]\, dt + \Sigma dW_P(t)$ where K_P is a constant $N \times N$ mean-reversion matrix, θ_P is a constant steady-state $N \times 1$ vector for $X(t)$, Σ is a constant $N \times N$ volatility matrix, and $W_P(t)$ is an $N \times 1$ vector of independent Brownian motions. Define the market prices of risk as $\Gamma(t) = \Sigma^{-1}[\gamma_0 + \gamma_1 X(t)]$, where γ_0 is a constant $N \times 1$ vector and γ_1 is an $N \times N$ matrix. Under the risk-neutral Q measure, the state variables follow the process $dX(t) = K_Q[\theta_Q - X(t)]\, dt + \Sigma dW_Q(t)$, where $dW_Q(t) = dW_P(t) + \Gamma(t)\, dt$, $K_Q = K_P + \gamma_1$, and $\theta_Q = (K_P + \gamma_1)^{-1}(K_P \theta_P - \gamma_0)$. Zero-coupon bond prices for the GATSM are $P(t,T) = A(t,T) + B(t,T)' X(t)$, where $B(t,T) = \left[\exp\left(-K_Q' \tau\right) - I\right]\left(K_Q'\right)^{-1} \rho_1$ with $\tau = T - t$ the time to maturity, and I the $N \times N$ identity matrix. The full expression for $A(t,T)$ is provided in Dai and Singleton (2002), but this article requires only the summary results that it has the functional form $A(\tau)$ and is required for the system to be arbitrage free.

From Heath, Jarrow and Morton (1992), instantaneous forward rates are defined as $f(t,T) = -\partial \log P(t,T)/\partial T$, and so:

$$f(t,T) = \left\{\exp\left(-K_Q' \tau\right)\left(K_Q'\right)^{-1} \rho_1\right\}' X(t) - \frac{\partial}{\partial \tau} A(\tau) \tag{1}$$

Now express K_Q' in eigensystem form; i.e $K_Q' = Z \Psi Z^{-1}$, where Z is the $N \times N$ non-singular matrix of eigenvectors each normalised to 1, and Ψ is the $N \times N$ diagonal matrix of unique and positive eigenvalues.[5] Hence, $\exp\left(-K_Q' \tau\right) = \exp\left(-Z \Psi Z^{-1} \tau\right) = Z \exp(-\Psi \tau) Z^{-1} = Z \Lambda Z^{-1}$, where $\Lambda =$diag$[\exp(\lambda_1 t), \ldots, \exp(\lambda_n t), \ldots, \exp(\lambda_N t)]$. The forward rates in equation 1 are then $f(t,T) = \left\{Z \Lambda Z^{-1} (K')^{-1} \rho_1\right\}' X(t) - \frac{\partial}{\partial \tau} A(\tau)$, which can be expressed equivalently as:

$$f(t,T) = \sum_{n=1}^{n_0} q_n(t) \exp(-\lambda_n \tau) + \sum_{n=n_0+1}^{N} q_n(t) \exp(-\lambda_n \tau) - \frac{\partial}{\partial \tau} A(\tau) \tag{2}$$

[3] In the notation of Dai and Singleton (2000), the specification is denoted $\mathrm{A}_0(N)$, and is the fully-Gaussian subset of the affine framework outlined in Duffie and Kan (1996). Being fully Gaussian, the results do not extend to Cox, Ingersoll and Ross (1985)/square-root dynamics. Indeed, it can readily be shown by example (available from the author) that affine models with square-root dynamics cannot "naturally" be reduced to the NS factor loadings as in section 3.

[4] Duffie and Kan (1996) notes that the state variables in an affine model can always be related back to economic factors (e.g technology, consumption, inflation, etc.) within a general equilibrium model, and Duffie and Singleton (1999) extends the latter result to financial factors (e.g default risk, liquidity risk, repo effects, etc.).

[5] This follows the standard assumptions of Duffie and Kan (1996) and Dai and Singleton (2002).

where the coefficients $q_n(t)$ represent the collection of coefficients associated with each unique $\exp(-\lambda_n\tau)$ term that arises from the full matrix multiplication of $\left\{Z\Lambda Z^{-1}(K')^{-1}\rho_1\right\}' X(t)$. Note that for convenience in the following section's example, but without loss of generality, it is assumed that the $\exp(-\lambda_n\tau)$ terms have been re-ordered from the smallest to the largest eigenvalue λ_n, and then grouped into the eigenvalues λ_1 to λ_{n_0} that are close to zero and the eigenvalues $\lambda_{n_{0+1}}$ to λ_N that are not close to zero.

3 The generic GATSM and the original NS model

From the exact expression of the generic GATSM forward rate curve in equation 2, three approximations are required to reproduce the original NS model. First, drop the arbitrage-free (hereafter AF) term $\frac{\partial}{\partial\tau}A(\tau)$. Second, for the first group of eigenvalues, $\lambda_n \simeq 0$ and the first term of the Taylor expansion is $\exp(-\lambda_n\tau) \simeq 1$. Third, for the second group of eigenvalues, express them relative to ϕ =median$(\lambda_{n_0+1},\ldots,\lambda_N)$,[6] so that $\lambda_n = \phi(1-\delta_n)$, and then $\exp(-\lambda_n\tau) = \exp(-\phi\tau)\exp(\delta_n\phi\tau)$. Now take the first-order Taylor approximation $\exp(\delta_n\phi\tau) \simeq 1+\delta_n\phi\tau$, and so $q_n(t)\exp(-\lambda_n\tau) \simeq q_n(t)\exp(-\phi\tau) + q_n(t)\,\delta_n\phi\tau\exp(\phi\tau)$. Substituting these results into 2 gives:

$$f(t,T) \simeq \sum_{n=1}^{n_0} q_n(t) + \sum_{n=n_0+1}^{N} q_n(t)\exp(-\phi\tau) + \sum_{n=n_0+1}^{N} q_n(t)\,\delta_n\lambda_n\tau\exp(\phi\tau) \tag{3}$$

This is precisely the functional form of the original NS model of the forward rate curve, i.e:

$$f(t,T) \simeq f_{\mathrm{NS}}(t,T) = L(t) + S(t)\exp(-\phi\tau) + C(t)\,\phi\tau\exp(-\phi\tau) \tag{4}$$

where 1, $\exp(-\phi\tau)$, and $\phi\tau\exp(-\phi\tau)$ are the NS forward rate factor loadings, and $L(t) = \sum_{n=1}^{n_0} q_n(t)$, $S(t) = \sum_{n=n_0+1}^{N} q_n(t)$, and $C(t) = \sum_{n=n_0+1}^{N} q_n(t)\,\delta_n\phi$ are the NS coefficients. The usual relationship $R_{\mathrm{NS}}(t,T) = \frac{1}{\tau}\int_0^\tau f_{\mathrm{NS}}(s)\,ds$ then produces the familiar form of the NS model.[7]

This shows that the original NS model represents the generic GATSM with no AF term (i.e $\frac{1}{\tau}\int_0^\tau \frac{\partial}{\partial s}A(s)\,ds = \frac{1}{\tau}A(\tau)$ that would have carried through from the generic GATSM forward rate curve), the persistent generic GATSM components approximated to zeroth order, and the non-persistent components approximated to first order. The appropriate AF term for $R_{\mathrm{NS}}(t,T)$ has been calculated in Christensen, Diebold and Rudebusch (2007), which provides a representation of $\frac{1}{\tau}A(\tau)$ from the generic GATSM, and creates the AFNS model with respect to the NS factor loadings.[8]

It is worthwhile at this point to discuss the practical implications of the approximations inherent in the original NS model relative to the GATSM. First, while the theoretical need for AF adjustment terms in NS models was originally established in Filipović (1999), their relevance in practice has recently been questioned; e.g see Coroneo, Nyholm and Vidova-Koleva (2008). The exposition in this article also makes it clear that the dynamics of the state variables and market prices of risk from the GATSM will be completely captured in the estimated coefficients of either the AFNS or NS model. Hence, time-series analysis based on NS or AFNS coefficients will in principle be identical to within a constant.

Regarding the approximations that produce the NS factor loadings from the generic GATSM, the approximations are "natural" in the sense that each additional term in the NS model corresponds

[6] Any other central measure of $(\lambda_{n_0+1},\ldots,\lambda_N)$ would suffice for the exposition in this article. In practice, ϕ is an estimated parameter.

[7] That is, the interest rate curve is $R_{\mathrm{NS}}(t,T) = L(t) + S(t)\left(\frac{1-\exp(-\phi\tau)}{\phi\tau}\right) + C(t)\left(\frac{1-\exp(-\phi\tau)}{\phi\tau} - \exp(-\phi\tau)\right)$, where $L(t)$, $S(t)$, and $C(t)$ are as already defined previously.

[8] The Christensen et al. (2007) AFNS model is a particular three-factor GATSM that by construction reproduces the three NS factor loadings with the NS coefficients as state variables. That is distinct from this article which shows how NS models represent the generic GATSM (and hence all GATSMs).

precisely to an additional term in the Taylor expansion of the generic GATSM. The approximations are also "transparent" in the sense that they will always represent the "true" GATSM to a known order regardless of (or even without knowledge of) the actual number of state variables and their interactions. It is worth noting the contrast here with the practical application of GATSMs, where necessary assumptions/approximations cloud the relationship between the GATSM and the "true" GATSM. For example, the application of GATSM models usually limits the number of state variables to just three,[9] often places prior constraints on parameters (e.g a diagonal covariance matrix), and commonly sets statistically-insignificant parameters from a first-pass estimation to zero to counter over-parametrisation.[10]

4 Classifying, applying, and interpreting NS models

By following the example in the previous section, specific NS models may be classified as various representations of the generic GATSM. The key aspects are the number of groups of eigenvalues (which determines the number of decay parameters), the degree of approximation within each group around the central eigenvalue (which determines the number of components containing each decay parameter), and whether the AF yield adjustment is included (which determines if the model is AF with respect to its factor loadings). The various permutations and extensions of each of those aspects can obviously generate a wide variety of potential NS models, but this section discusses just the range of NS models already in use and two interesting extensions that add just one component to the original NS model.

The $[L(t), S_1(t), C_1(t), S_2(t), C_2(t), \phi_1, \phi_2]$ Christensen et al. (2008)/NS model is most comprehensive version of the NS class to date. It effectively represents the persistent GATSM components to zeroth order, the non-persistent components with two groupings of non-zero eigenvalues (moderate $\phi_1 = \text{median}(\lambda_{n_0+1}, \ldots, \lambda_{n_1})$ and high $\phi_2 = \text{median}(\lambda_{n_1+1}, \ldots, \lambda_N)$) each to first order, and includes an AF term to ensure the model is AF with respect to the five factor loadings. The $[L(t), S_1(t), C_1(t), C_2(t), \phi_1, \phi_2]$ Svensson (1995)/NS model omits the yield adjustment term and $S_2(t)$, the first term of the second approximation.

At the other extreme, the $[L(t), S(t), \phi]$ Diebold et al. (2008)/NS model is the simplest representation of the generic GATSM; it ignores the AF adjustment and represents both the persistent and non-persistent components of the generic GATSM to zeroth order. In terms of the degree of approximation then, the Diebold et al. (2008)/NS model is "balanced". A balanced version of the original NS model would approximate the persistent GATSM components to first order to match the first-order approximation inherent in the Slope and Curvature components.[11] Alternatively, adding another Slope component to the original NS model, i.e $[L(t), S_1(t), S_2(t), \phi_1, \phi_2]$, would also be a balanced model, but approximated to zeroth order.

Choosing which specific NS model to apply will continue to be a largely an empirical matter; i.e trading off parsimony against goodness of fit to the yield curve data. However, the classification above suggests a systematic approach to introducing or omitting new terms. For example, the Svensson (1995)/NS model should be avoided because the second Curvature term $C_2(t)$ cannot by itself represent a natural first-order approximation of the GATSM. Either the second Slope term $S_2(t)$ should also be added (creating the non-AF analog of the Christensen et al. (2008)/NS model), or the second Curvature term $C_2(t)$ dropped (recreating the original NS model). Another aspect is more subtle: from the strict perspective of maintaining a foundation within the generic GATSM, NS models should be applied with a constant decay parameter(s) ϕ. Allowing for changes each time the model is re-estimated means that the NS coefficients lose their correspondence to the state

[9] Duffee (2002) notes that "general three-factor affine models are already computationally difficult to estimate owing to the number of parameters. Adding another factor would make this investigation impractical."

[10] Examples are Dai and Singleton (2002) and Duffee (2002).

[11] Specifically, $\exp(-\lambda_n \tau) \simeq 1 - \lambda_n \tau$, and so $\sum_{n=1}^{n_0} q_n(t) \exp(-\lambda_n \tau) \simeq \sum_{n=1}^{n_0} q_n(t) - \sum_{n=1}^{n_0} q_n(t) \lambda_n \tau$. Hence, the Level component would be adjusted by a small negative amount that becomes increasingly negative by time-to-maturity.

variables in the GATSM, and the decay parameter(s) loses its interpretation as a central measure of mean-reversion in the GATSM.[12]

On the face of it, the interpretation of NS models from an economic and financial perspective might seem an impossible task, given that the NS coefficients are reduced forms of Taylor approximations to the underlying GATSM. However, in aggregate the estimated Level and non-Level coefficients of the NS model respectively reflect the persistent and non-persistent components of the generic GATSM, and so the context of the application can provide an indication of the economic or financial variables that the Level and non-Level components represent. For example, one might expect the Level coefficient from the estimation of a nominal government yield curve to correspond to inflation/expected inflation (persistent economic variables), and the Slope and Curvature coefficients to correspond with output growth/expected output growth (cyclical economic variables). Such relationships have already been established empirically in Diebold, Rudebusch and Aruoba (2006) using the original NS model.

The interpretation of NS models is enhanced when considering spreads between yield curves, given that the influence/s underlying those spreads are usually apparent from the context. For example, if the inflation-indexed and nominal yield curves are both GATSMs, then the spread between those curves will also be a GATSM, but only in the state variables that differ between the two curves. Hence, applying an NS model directly to the differences between the inflation-indexed and nominal yield curve would represent a GATSM in expected inflation and liquidity and the term premia associated with the uncertainty and market prices of risk in those factors.

5 Conclusion

This article shows how specific models of the NS class may be classified as various representations of the generic GATSM. That provides a compelling case for applying NS models as standard tools for yield curve analysis in economics and finance; the user will get the same results in practice, but without the complexity associated with the practical application of GATSMs.

References

Bank for International Settlements (2005), 'Zero-coupon yield curves: technical documentation', *BIS Papers* **25**.

Christensen, J., Diebold, F. and Rudebusch, G. (2007), 'The affine arbitrage-free class of Nelson-Siegel term structure models', *NBER Working Paper Series* **13611**.

Christensen, J., Diebold, F. and Rudebusch, G. (2008), 'An arbitrage-free generalized Nelson-Siegel term structure model', *Journal of Econometrics (forthcoming)* .

Coroneo, L., Nyholm, K. and Vidova-Koleva, R. (2008), 'How arbitrage-free is the Nelson-Siegel model?', *Working Paper, European Central Bank* **874**.

Cox, J., Ingersoll, J. and Ross, S. (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**, 385–407.

Dai, Q. and Singleton, K. (2000), 'Specification analysis of affine term structure models', *Journal of Finance* **55(5)**, 1943–1978.

Dai, Q. and Singleton, K. (2002), 'Expectation puzzles, time-varying risk premia, and affine models of the term structure', *Journal of Financial Econometrics* **63**, 415–441.

[12] While it would be tempting to interpret time variation in ϕ as representing time variation in the mean-reversion matrix $K_Q = K_P + \gamma_1$, a GATSM that formally allowed for such flexible time variation would result in more complex factor loadings that could no longer "naturally" be reduced to the NS factor loadings as in section 3.

Diebold, F., Li, C. and Yue, V. (2008), 'Global yield curve dynamics and interactions: a dynamic Nelson-Siegel approach', *Journal of Econometrics* **146**, 351–363.

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

Duffee, G. (2002), 'Term premia and interest rate forecasts in affine models', *Journal of Finance* **57(1)**, 405–443.

Duffie, D. and Kan, R. (1996), 'A yield-factor model of interest rates', *Mathematical Finance* **6(4)**, 379–406.

Duffie, D. and Singleton, K. (1999), 'Modelling term structures of defaultable bonds', *Review of Financial Studies* **12(4)**, 687–720.

Filipović, D. (1999), 'A note on the Nelson-Siegel family', *Mathematical Finance* **9(4)**, 349–359.

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

Heath, D., Jarrow, R. and Morton, A. (1992), 'Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation', *Econometrica* **60(1)**, 77–106.

Nelson, C. and Siegel, A. (1987), 'Parsimonious modelling of yield curves', *Journal of Business* **October**, 473–489.

Svensson, L. (1995), 'Estimating forward interest rates with the extended Nelson and Siegel model', *Quarterly Review, Sveriges Riksbank* **1995(3)**, 13–26.

Yu, W. and Salyards, D. (2009), 'Parsimonious modeling and forecasting of corporate bond yields', *Journal of Forecasting* **28**, 73–88.