Analyzing the Term Structure of Interest Rates using the Dynamic Nelson-Siegel Model with Time-Varying Parameters

Siem Jan Koopman[a,c] *Max I.P. Mallee*[a] *Michel van der Wel*[b,c]

[a] Department of Econometrics, VU University Amsterdam

[b] Department of Finance, VU University Amsterdam

[c] Tinbergen Institute, Amsterdam

Some keywords: Generalized autoregressive conditional heteroskedasticity model; Extended Kalman filter; Time-varying volatility; Yield curve.

JEL classification: C32, C51, E43.

Acknowledgments

We are grateful to the Editor, an Associate Editor and two referees for their insightful comments. The paper has improved significantly as a result. We further would like to thank Francis X. Diebold, Dick van Dijk and Michiel de Pooter for their comments on an earlier version of this paper. Particular thanks are due to our colleague Charles S. Bos for his computational support and assistance. We have benefited from the comments of participants in seminar presentations at the Erasmus University Rotterdam, the University of Pennsylvania, the NAKE 2007 day in Utrecht, the Tinbergen Institute Amsterdam and the Center for Research in Econometric Analysis of Time Series (CREATES) in Aarhus. Finally, we would like to thank Francis X. Diebold for making the dataset available on his website: `www.ssc.upenn.edu/~fdiebold/`. More details of our estimation results generated for this paper are available from the authors upon request. All errors are our own.

Analyzing the Term Structure of Interest Rates using the Dynamic Nelson-Siegel Model with Time-Varying Parameters

Abstract

In this paper we introduce time-varying parameters in the dynamic Nelson-Siegel yield curve model for the simultaneous analysis and forecasting of interest rates of different maturities, known as the term structure. The Nelson-Siegel model has been recently reformulated as a dynamic factor model where the latent factors are interpreted as the level, slope and curvature of the term structure. The factors are modeled jointly as a vector autoregressive process. We propose to extend this framework in two directions. First, the factor loadings in the Nelson-Siegel yield model depend on a single loading parameter. We allow this parameter to be time-varying by treating it as the fourth latent factor that is modeled jointly with the other factors in the vector autoregressive process. Second, we investigate in detail whether the overall volatility in interest rates is constant over time. For this purpose, we introduce a common volatility component that is specified as a GARCH (generalized autoregressive conditional heteroskedasticity) process. The common volatility component is scaled separately for each maturity by an unknown coefficient. We further investigate whether the innovations of the factors are also subject to a common volatility component. Based on a data-set of yield curves that is analyzed by others, we present empirical evidence of considerable increases in model fit when time-varying loadings and volatilities in the dynamic Nelson-Siegel are introduced.

1 Introduction

Fitting and predicting time-series of a cross-section of yields has proven to be a challenging task. As with many topics in empirical economic analysis there is the trade-off between the goodness of fit that is obtained by employing statistical models without a reference to economic theory, and the lack of fit by economic models that do provide a basis for the underlying economic theory.

For many decades work on the term structure of interest rates has mainly been theoretical in nature. In the early years work focused on the class of affine term structure models, see

Vasicek (1977) and Cox, Ingersoll, and Ross (1985). Duffie and Kan (1996) generalized the literature and Dai and Singleton (2000) characterized the set of admissable and identifiable models. Another class of models focused on fitting the term structure at a given point in time to ensure no arbitrage opportunities exist, see Hull and White (1990) and Heath, Jarrow, and Morton (1992). It has been shown that the forecasts obtained using the first class of models do not outperform the random walk forecasts, see for example Duffee (2002). The second class of models are not appropriate for forecasting given its focus on the cross-section dimension of yields without a reference to the time series dimension. Time series models aim to describe the dynamical properties and are therefore more suited for forecasting. This may partly explain the renewed interest in statistical time series models for yield curves.

The papers of Diebold and Li (2006, DL) and Diebold, Rudebusch, and Aruoba (2006, DRA) have shifted attention back to the Nelson and Siegel (1987) model. DL and DRA introduce the dynamic Nelson-Siegel model as a statistical three factor model to describe the yield curve over time. The three factors represent level, slope and curvature of the yield curve and thus carry some level of economical interpretation. More importantly, DL and DRA show that the model-based forecasts outperform many other models including standard time series models such as vector autoregressive models and dynamic error-correction models. In DRA, the Nelson-Siegel framework is extended to include non-latent factors such as inflation. Further they frame the Nelson-Siegel model into a state space model where the three factors are treated as unobserved processes and modeled by vector autoregressive processes. A wide range of statistical methods associated with the state space model can be exploited for maximum likelihood estimation and signal extraction, see Durbin and Koopman (2001). We will follow this approach in which the state space representation of the Nelson and Siegel (1987) model plays a central role.

Parameter estimation in DL and DRA relies on two simplifying assumptions. First, the factor loadings in the Nelson-Siegel model depend on a single loading parameter. To enable the estimation of time-varying latent factors in a linear setting, the factor loadings are kept constant over time for each maturity. In the original Nelson and Siegel (1987) article, the loading parameter is estimated based on the yield (cross-section) for each time period. In DL, the loading parameter is restricted as constant to keep the factor loadings constant. Second, volatility is kept constant for each maturity and over the full sample period.

We contribute to the literature by introducing time-varying factor loadings and time-varying volatility in the dynamic Nelson-Siegel model. The loading parameter determines all factor loadings in the Nelson-Siegel model. When we allow the loading parameter to be time-varying, all loadings will be time-varying. We consider the loading parameter as a stochastically time-varying latent process and we treat it as the fourth factor. The latent factors level, slope and curvature together with the time-varying loading parameter are then modeled jointly as a vector autoregressive process. The loading parameter is not a linear function of the observation vector and therefore we obtain a nonlinear dynamic model. We will show that the nonlinear features in the dynamic Nelson-Siegel model can be treated using extended Kalman filter methods. Next, we introduce time-varying volatility by specifying the common variance as the well-known generalized autoregressive conditional heteroskedasticity (GARCH) process, see Bollerslev (1986). In empirical work it is found that during high volatility periods, the yields for all maturities are highly volatile although some maturities are more volatile than others. The time-varying volatility is introduced for different components of the model. We finally carry out formal test procedures and we present misspecification diagnostics to assess the most appropriate time-varying specifications in the dynamic Nelson-Siegel model.

The introduction of time-varying parameters may also shed some light on more recent developments in the term structure literature. The dynamic Nelson-Siegel model does not rely on theoretical concepts such as the absence of arbitrage, see also the discussion in Ang and Piazzesi (2003). Recently, Christensen, Diebold, and Rudebusch (2007) have modified the Nelson-Siegel framework to impose the arbitrage-free condition. As a result, a new class of affine dynamic term structure models is defined. An important condition for the risk-free rate to exist in this framework is that loadings are constant over time. This condition may be validated by allowing the factor loadings to be time-varying as we do in this paper. Also in the work of Diebold, Li, and Yue (2007) on the global yield curve, constant factors are an important condition.

There are a number of papers that extend the work of DL and DRA for the Nelson-Siegel model. Bianchi, Mumtaz, and Surico (2006) allow for time-varying variances for the latent factors level, slope and curvature. In this specification, it is implied that the factor loadings for the term-structure are also appropriate weights for the volatility in the term-structure. This appears to be a strong assumption that needs to be validated. We therefore introduce for each yield in the observation equation, a different factor loading for the common disturbance factor that is subject to a GARCH process. We then construct a test whether the common volatility factor loadings are a linear combination of the loadings for the factors level, slope and curvature. Yu and Zivot (2007) extend the Nelson-Siegel framework by including corporate bonds. De Pooter (2007) examines the dynamic NS model that is extended by additional factors. It is shown that such extensions can improve both the in-sample fit and the post-sample forecasting performance. Without adopting the Nelson-Siegel framework, Bowsher and Meeks (2008) introduce a 5-factor model where spline functions are used to model the yield curve and where the knots for these splines act as factors. While their approach allows for a more flexible yield curve some economic intuition of the factors is lost. Moreover, also in this framework, volatility is kept fixed over time.

The remainder of the paper is organized as follows. Section 2 describes the baseline dynamic Nelson-Siegel model and Section 3 discusses our new extensions. In Section 4 we present, discuss and compare estimation results for different model specifications. Section 5 concludes.

2 The dynamic Nelson-Siegel model

In this section we introduce the latent factor model that Nelson and Siegel (1987) have developed for the yield curve. We focus on the model that is slightly adjusted in terms of factorization by Diebold and Li (2006) and is extended here to allow for time-varying parameters. We further discuss the state space approach for this initial extension of the model.

2.1 The Nelson-Siegel model

Interest rates are denoted by $y_t(\tau)$ at time t and maturity τ. For a given time t, the yield curve $\theta_t(\tau)$ is some smooth function representing the interest rates (yields) as a function of maturity τ. A parsimonious functional description of the yield curve is proposed by Nelson and Siegel (1987). The Nelson-Siegel formulation of the yield is modified by Diebold and Li (2006) to lower the coherence between the components of the yield curve.

The Diebold and Li (2006) formulation is given by

$$\theta_t(\tau) = \theta(\tau; \lambda, \beta_t) = \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda \tau}}{\lambda \tau} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda \tau}}{\lambda \tau} - e^{-\lambda \tau} \right), \tag{1}$$

where $\beta_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})'$, for given time t, maturity τ and fixed coefficient λ that determines the exponential decay of the second and third component in (1).

The shape and form of the yield curve is determined by the three components and their associated weights in β_t. The first component takes the value 1 (constant) and can therefore be interpreted as the overall level that influences equally the short and long term interest rates. The second component converges to one as $\tau \downarrow 0$ and converges to zero as $\tau \to \infty$ for a given t. Hence this component mostly influences short-term interest rates. The third component converges to zero as $\tau \downarrow 0$ and as $\tau \to \infty$ but is concave in τ, for a given t. This component is therefore associated with medium-term interest rates.

Since the first component is the only one that equals one as $\tau \to \infty$, its corresponding β_{1t} coefficient is usually linked with the long-term interest rate. By defining the slope of the yield curve as $\theta_t(\infty) - \theta_t(0)$, it is easy to verify that the slope converges to $-\beta_{2t}$ for a given t. Finally, the shape of the yield can be defined by $[\theta_t(\tau^*) - \theta_t(0)] - [\theta_t(\infty) - \theta_t(\tau^*)]$ for a medium maturation τ^*, say, two years, and for a given t. It can be shown that this shape approximately equals β_{3t}.

In case we observe a series of interest rates $y_t(\tau_i)$ for a set of N different maturities $\tau_1 < \ldots < \tau_N$ available at a given time t, we can estimate the yield curve by the simple regression model

$$\begin{aligned} y_t(\tau_i) &= \theta_t(\tau_i) + \varepsilon_{it} \\ &= \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda \tau_i}}{\lambda \tau_i} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda \tau_i}}{\lambda \tau_i} - e^{-\lambda \tau_i} \right) + \varepsilon_{it}, \end{aligned} \tag{2}$$

for $i = 1, \ldots, N$. The disturbances $\varepsilon_{1t}, \ldots, \varepsilon_{Nt}$ are assumed to be independent with mean zero and constant variance σ_t^2 for a given time t. The least squares method provides estimates for the β_{jt} coefficients $j = 1, 2, 3$. These cross-section estimates can be obtained as long as sufficient interest rates for different maturities are available at time t.

The series of regression estimates for β_t, for all time periods $t = 1, \ldots, T$, appear to be strongly correlated over time. In other words, the coefficients are forecastable and hence the Nelson-Siegel framework can be used for forecasting in this way. This has been recognized by Diebold and Li (2006) who implemented the following two-step procedure: first, estimate the β_t by cross-section least squares for each t; second, treat these estimates as three time series and apply time series methods for forecasting β_t and hence the yield curve $\theta(\tau; \lambda, \beta_t)$.

Diebold and Li (2006) compare their two-step forecasts with those from univariate and multivariate time series methods. The different methods produce similar results. Nevertheless, the two-step forecasting approach does better than forecasting the different interest rates series directly, especially for the longer maturities.

2.2 The dynamics of the latent factors

Diebold, Rudebusch, and Aruoba (2006) go a step further by recognizing that the Nelson-Siegel framework can be represented as a state space model when treating β_t as a latent vector. For this purpose, the regression equation (2) is rewritten by

$$y_t = \Lambda(\lambda)\beta_t + \varepsilon_t, \qquad \varepsilon_t \sim NID(0, \Sigma_\varepsilon), \qquad t = 1, \ldots, T, \tag{3}$$

with observation vector $y_t = [y_t(\tau_1), \ldots, y_t(\tau_N)]'$, disturbance vector $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$ and $N \times 3$ factor loading matrix $\Lambda(\lambda)$ where its (i, j) element is given by

$$\Lambda_{ij}(\lambda) = \begin{cases} 1, & j = 1, \\ \left(1 - e^{-\lambda \cdot \tau_i}\right) \ / \ \lambda \cdot \tau_i, & j = 2, \\ \left(1 - e^{-\lambda \cdot \tau_i} - \lambda \cdot \tau_i e^{-\lambda \cdot \tau_i}\right) \ / \ \lambda \cdot \tau_i, & j = 3. \end{cases}$$

The time series process for the 3×1 vector β_t can be modeled by the vector autoregressive (VAR) process

$$\beta_{t+1} = (I - \Phi)\mu + \Phi\beta_t + \eta_t, \qquad \eta_t \sim NID(0, \Sigma_\eta), \tag{4}$$

for $t = 1, \ldots, n$, with mean vector μ and initial condition $\beta_1 \sim N(\mu, \Sigma_\beta)$ where coefficient matrix Φ and variance matrix Σ_β are chosen such that $\Sigma_\beta - \Phi\Sigma_\beta\Phi' = \Sigma_\eta$ and stationarity of the VAR process must be ensured, see Ansley and Kohn (1986) for an effective reparameterisation. We refer to model (3) and (4) as the dynamic Nelson-Siegel (DNS) model.

2.3 Estimation based on the Kalman filter

We consider the DNS model (3) and (4) as a linear Gaussian state space model. The vector of unobserved factors β_t is the state vector and can be estimated conditional on the past and concurrent observations $y_1, \ldots, y_t$ via the Kalman filter. Define $b_{t|s}$ as the minimum mean square linear estimator (MMSLE) of β_t given $y_1, \ldots, y_s$ with mean square error (MSE) matrix $B_{t|s}$, for $s = t-1, t$. For given values of $b_{t|t-1}$ and $B_{t|t-1}$, the Kalman filter first computes $b_{t|t}$ and $B_{t|t}$, when observation y_t is available, using the filtering step

$$b_{t|t} = b_{t|t-1} + B_{t|t-1}\Lambda(\lambda)' F_t^{-1} v_t, \qquad B_{t|t} = B_{t|t-1} - B_{t|t-1}\Lambda(\lambda)' F_t^{-1}\Lambda(\lambda) B_{t|t-1}, \tag{5}$$

where $v_t = y_t - \Lambda(\lambda) b_{t|t-1}$ is the prediction error vector and $F_t = \Lambda(\lambda) B_{t|t-1}\Lambda(\lambda)' + \Sigma_\varepsilon$ is the prediction error variance matrix. The MMSLE of the state vector for the next period $t+1$, conditional on $y_1, \ldots, y_t$, is given by the prediction step

$$b_{t+1|t} = (I - \Phi)\mu + \Phi b_{t|t}, \qquad B_{t+1|t} = \Phi B_{t|t}\Phi' + \Sigma_\eta. \tag{6}$$

For a given time series $y_1, \ldots, y_T$, the Kalman filter computations are carried out recursively for $t = 1, \ldots, T$ with initializations $b_{1|0} = \mu$ and $B_{1|0} = \Sigma_\beta$ where Σ_β is defined below (4). The parameters in the VAR coefficient matrix Φ, variance matrices Σ_η and Σ_ε together with μ and λ are treated as unknown coefficients which are collected in the parameter vector ψ. Estimation of ψ is based on the numerical maximization of the loglikelihood function that is constructed via the prediction error decomposition and given by

$$\ell(\psi) = -\frac{NT}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log|F_t| - \frac{1}{2}\sum_{t=1}^{T} v_t' F_t^{-1} v_t. \tag{7}$$

As a result, $\ell(\psi)$ can be evaluated by the Kalman filter for a given value of ψ. A quasi-Newton optimization method is employed for the purpose of maximization based on the numerical evaluation of the score function. We have implemented the necessary calculations in the matrix programming language `Ox` of Doornik (2001) with the use of the `SsfPack` state space functions developed by Koopman, Shephard, and Doornik (1999). A textbook treatment of Kalman filter methods is given by Durbin and Koopman (2001).

The state space framework allows for different dynamic processes for vector β_t in the dynamic Nelson-Siegel model. Also, variance matrices Σ_ε and Σ_η can be full or diagonal. Diebold, Rudebusch, and Aruoba (2006) assume that Σ_ε is diagonal so that the equations for the different yield maturities are uncorrelated, given β_t. This assumption is often used to reduce the number of coefficients and to obtain computational tractability.

3 DNS model with time-varying parameters

In this section we extend the DNS model by treating the loading parameter as a stochastically time-varying latent factor and by introducing time-varying volatility in the variance specification of the disturbances. The extensions introduce nonlinearities in the model that we will handle by the extended Kalman filter discussed below.

3.1 Time-varying loading parameter

In the DNS model, the loading parameter λ determines the shape of the yield curve. In the earlier studies, the default is to pre-fix a value for λ without estimation. For example, Diebold and Li (2006) fix λ at 0.0609 while Diebold, Rudebusch, and Aruoba (2006) estimate λ to be 0.077. Yu and Zivot (2007) adopt these values for λ in their empirical study concerning corporate bonds. They argue that the loadings $\Lambda_{ij}(\lambda)$ are not very sensitive to different values of λ as can be illustrated graphically. Hence they argue that λ can be fixed such that it maximizes the loading on the curvature component at some medium term (that is, 30 months for $\lambda = 0.0609$ and 23.3 months for $\lambda = 0.077$).

Here we emphasize that the estimation of λ is straightforward in a state space framework as it can be included in the parameter vector ψ, see Section 2.3. Keeping λ fixed over the full sample period may be too restrictive as the data usually spans over a long time period. In particular, the maturity at which the curvature factor β_{3t} is maximized and the speed of decay of the slope factor β_{2t} depend only on λ and are fixed as a result. However, these characteristics of the yield curve may have changed over time. The importance of λ and its constancy over time is also discussed in Christensen, Diebold, and Rudebusch (2007) where an arbitrage-free version of the Nelson-Siegel framework is proposed. Given the importance of λ, we study its role in more detail by considering possible changes of λ over time.

We propose to treat the time-varying loading parameter λ_t as the latent fourth factor β_{4t} and to include it into the vector β_t below (1). The new 4×1 vector β_t is then modeled by the VAR process (4). The loading matrix $\Lambda(\lambda)$ in the observation equation (3) is then replaced by $\Lambda(\lambda_t)$ with $\lambda_t = \beta_{4t}$. The new observation equation is nonlinear in β_t as we obtain $y_t = \Lambda(\beta_{4t}) \cdot (\beta_{1t}, \beta_{2t}, \beta_{3t})' + \varepsilon_t$. This specification is of particular interest since it allows dynamic interactions between changes in cross-sectional (or cross-maturity) dependence, through β_{4t}, and time series dependence of the yields, through β_{1t}, β_{2t} and β_{3t}. The resulting model with time-varying loadings will be referred to as the DNS–TVL model. Other dynamic processes for λ_t can be considered including a random walk process.

3.2 Time-varying volatility

Another key aspect in the analysis of the term structure is the recognition that interest rates are the result of trading at financial markets. The volatility in the series may therefore have changed over time as well. In most empirical work on the yield curve, monthly time series of interest rates are analyzed under the assumption that the volatility in the time series is constant over time. A few exceptions are Engle, Ng, and Rothschild (1990) and Bianchi, Mumtaz, and Surico (2006). However, investigating time-varying volatility in the context of the DNS model is a novelty. Although the changes in the volatilities for the different maturities have different intensities, they appear to occur at the same time. For this reason, we focus mainly on a common pattern of time-varying volatility in interest rates.

Here we modify the DNS model by introducing time-varying variance matrices via a common volatility component that is modeled by a GARCH process. We adopt the common GARCH specification, or the one-factor GARCH model, of Harvey, Ruiz, and Sentana (1992) to introduce a time-varying variance for the disturbances in the observation equation. In particular, we consider the decomposition of the disturbance vector ε_t given by

$$\varepsilon_t = \Gamma \varepsilon_t^* + \varepsilon_t^+, \qquad t = 1, \ldots, T,$$

where Γ is redefined here as a $N \times 1$ loading vector, ε_t^* is a scalar disturbance and ε_t^+ is a $N \times 1$ disturbance vector. For identification purposes, vector Γ can be normalized such that $\Gamma'\Gamma = 1$. The disturbance components are mutually independent of each other and their distributions are given by

$$\varepsilon_t^* \sim NID(0, h_t), \qquad \varepsilon_t^+ \sim NID(0, \Sigma_\varepsilon^+), \qquad t = 1, \ldots, T, \tag{8}$$

where Σ_ε^+ is typically , but not necessarily, a diagonal matrix and where variance h_t is specified as the GARCH process developed by Bollerslev (1986). In particular, we have

$$h_{t+1} = \gamma_0 + \gamma_1 {\varepsilon_t^*}^2 + \gamma_2 h_t, \qquad t = 1, \ldots, T, \tag{9}$$

with unknown coefficients $\gamma_0 > 0$, $0 < \gamma_1 < 1$ and $0 < \gamma_2 < 1$ and $h_1 = \gamma_0(1 - \gamma_1 - \gamma_2)^{-1}$. Lags of ε_t^* and h_t can also be included in the specification (9). As a result, the variance matrix of ε_t has become time-varying and is given by

$$\Sigma_\varepsilon(h_t) = h_t \Gamma\Gamma' + \Sigma_\varepsilon^+, \qquad t = 1, \ldots, T, \tag{10}$$

where $\Sigma_\varepsilon(h_t)$ is a full variance matrix but its time-variation depends on the common and univariate GARCH process (9). In this specification, the normalization constraint $\Gamma'\Gamma = 1$ can be replaced by fixing γ_0 to a known constant. The (unconditional) time-varying variance matrix of y_t is $\Lambda(\lambda)\Sigma_\beta\Lambda(\lambda)' + \Sigma_\varepsilon(h_t)$ where Σ_β is the solution of $\Sigma_\beta - \Phi\Sigma_\beta\Phi = \Sigma_\eta$. The unknown coefficients for the GARCH specification are collected in the parameter vector $\gamma = (\gamma_1, \gamma_2, \Gamma')'$. We treat γ_0 as a known constant and refer to this model as DNS–GARCH.

In the same framework, we can consider a GARCH specification for Σ_η in (4) based on the decomposition $\eta_t = \Gamma_\eta \eta_t^* + \eta_t^+$ with $\eta_t^* \sim NID(0, g_t)$ and $\Sigma_\eta = \Sigma_\eta(g_t)$ as in (10). When we focus on this decomposition only, the variance matrix of y_t is given by $\Lambda(\lambda)\Sigma_\beta(g_t)\Lambda(\lambda)' + \Sigma_\varepsilon$. In this setting, the variance structure of y_t subject to volatility depends on matrix $\Lambda(\lambda)$. We regard this specification as a restriction compared to (10) where the volatility variance structure is determined by Γ. We can construct a likelihood-ratio statistic for this restriction. Another hypothesis can be formulated by considering $y_t = \Lambda(\lambda)\beta_t + \Gamma\varepsilon_t^* + \varepsilon_t^+$ without GARCH for Σ_η. The hypothesis of interest is $\Gamma = \Lambda(\lambda)w$ where w is a 3×1 vector of unknown coefficients. The null model is then given by $y_t = \Lambda(\lambda)\left(\beta_t + w\varepsilon_t^*\right) + \varepsilon_t^+$. In this case, the variance matrix of y_t has become $\Lambda(\lambda)\left[\Sigma_\beta + h_t w w'\right]\Lambda(\lambda)' + \Sigma_\varepsilon$ where the variance structure subject to volatility also depends on the factor loadings in $\Lambda(\lambda)$.

3.3 Estimation based on the extended Kalman filter

In case loading parameter λ_t and variance h_t (or g_t) are specified as an autoregressive process and a GARCH process, respectively, we cannot determine λ_t and h_t a-priori. In particular, λ_t depends on the Nelson-Siegel latent factors level, slope and curvature ($\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$, respectively) while the time-varying variance h_t in (9) depends deterministically on past values of the unobserved disturbance term ε_t^*. Therefore, we treat $\lambda_t = \beta_{4,t}$ and ε_t^* as latent variables of interest and place them in the state vector α_t. A nonlinear state space model can be designed for the DNS model with time-varying parameters based on this state vector. The nonlinear observation equation is given by

$$y_t = Z_t(\alpha_t) + \varepsilon_t^+, \qquad \varepsilon_t^+ \sim NID\{0, \Sigma_\varepsilon^+\}, \qquad t = 1, \dots, T, \tag{11}$$

with $\alpha_t = (\beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \beta_{4,t}, \varepsilon_t^*)' = (\beta_t', \varepsilon_t^*)'$ and where $Z_t(\alpha_t)$ is the $N \times 1$ vector function

$$Z_t(\alpha_t) = \Lambda(\lambda_t)(\beta_{1,t}, \beta_{2,t}\beta_{3,t})' + \Gamma\varepsilon_t^*, \qquad \text{with } \lambda_t = \beta_{4,t}, \qquad t = 1, \dots, T. \tag{12}$$

The state equation is given by

$$\alpha_{t+1} = c + \begin{bmatrix} \Phi & 0 \\ 0 & 0 \end{bmatrix} \alpha_t + \begin{pmatrix} \eta_t \\ \varepsilon_{t+1}^* \end{pmatrix}, \qquad \begin{pmatrix} \eta_t \\ \varepsilon_{t+1}^* \end{pmatrix} \sim NID\left(0, \begin{bmatrix} \Sigma_\eta & 0 \\ 0 & h_{t+1} \end{bmatrix}\right), \tag{13}$$

for $t = 1, \dots, T$ and $c = [\mu'(I - \Phi)', 0]'$. Since h_{t+1} in (9) is a function of the unobserved value ε_t^* and its past values, we will not be able to compute the necessary value of h_{t+1} at time t. A solution is to replace h_{t+1} by its estimate based on observations $y_1, \dots, y_t$, that is

$$\hat{h}_{t+1|t} = \gamma_0 + \gamma_1 e_t^2 + \gamma_2 \hat{h}_{t|t-1}, \qquad t = 1, \dots, T,$$

where e_t is an estimate of ε_t^* based on $y_1, \dots, y_t$ and obtained from the filtering step of the Kalman filter applied to the model (11) and (13). Past values of $\hat{h}_{t|t-1}$ can be stored outside the model and the variance h_{t+1} in (13) is replaced by $\hat{h}_{t+1|t}$ for the prediction step of the Kalman filter. As a result, the state estimates are sub-optimal, they are not MMSLE. A more detailed discussion of this approach is provided by Harvey, Ruiz, and Sentana (1992).

The Kalman filter method only applies to models that are linear in the state vector. The observation equation (11) is clearly nonlinear in α_t. Exact estimation procedures for nonlinear models require a major computational effort. We prefer to preserve the elegance of the Kalman filter. For this purpose, we locally linearize the nonlinear function $Z_t(\alpha_t)$ at $\alpha_t = a_{t|t-1}$ where $a_{t|t-1}$ is an estimate of α_t based on the past observations $y_1, \ldots, y_{t-1}$. We obtain the linearized model

$$y_t = Z_t(a_{t|t-1}) + \dot{Z}_t \cdot (\alpha_t - a_{t|t-1}) + \varepsilon_t^+ = d_t + \dot{Z}_t \alpha_t + \varepsilon_t^+, \qquad t = 1, \ldots, T,$$

where $d_t = Z_t(a_{t|t-1}) - \dot{Z}_t a_{t|t-1}$ and $\dot{Z}_t = \partial Z_t(\alpha_t) \,/\, \partial \alpha_t' \,|_{\alpha_t = a_{t|t-1}} = (\dot{z}_{1t}', \ldots, \dot{z}_{Nt}')'$ with

$$\dot{z}_{it} = \Big[1 \,,\, \Lambda_{i2}(a_{4,t|t-1}) \,,\, \Lambda_{i3}(a_{4,t|t-1}) \,,\, a_{2,t|t-1}\dot{\Lambda}_{i2}(a_{4,t|t-1}) + a_{3,t|t-1}\dot{\Lambda}_{i3}(a_{4,t|t-1}) \,,\, \Gamma_i \Big] \,,$$

for which the loading element $\Lambda_{ij}(\lambda)$ is given below (3), $\dot{\Lambda}_{ij}(x) = \partial \Lambda_{ij}(\lambda) \,/\, \partial \lambda|_{\lambda = x}$, $a_{k,t|t-1}$ is the kth element of vector $a_{t|t-1}$ and Γ_i is the ith element of vector Γ, for $i = 1, \ldots, N$, $j = 2, 3$ and $k = 2, 3, 4$. Given an estimate $a_{t|t-1}$ and an approximate MSE matrix $A_{t|t-1}$ for $a_{t|t-1}$, the filtering step is given by

$$a_{t|t} = a_{t|t-1} + A_{t|t-1}\dot{Z}_t' F_t^{-1} v_t, \qquad A_{t|t} = A_{t|t-1} - A_{t|t-1}\dot{Z}_t' F_t^{-1} \dot{Z}_t A_{t|t-1}, \tag{14}$$

with $v_t = y_t - d_t - \dot{Z}_t a_{t|t-1} = y_t - Z_t(a_{t|t-1})$ and $F_t = \dot{Z}_t A_{t|t-1} \dot{Z}_t' + \Sigma_\varepsilon^+$. We define $a_{t|t}$ as a sub-optimal estimate of α_t based on observations $y_1, \ldots, y_t$ and $A_{t|t}$ as its approximate MSE matrix. The prediction step is similar to (6) but then based on the state equation (13).

The estimates $a_{t|t-1}$ and $a_{t|t}$ are sub-optimal due to the replacement of h_{t+1} in (13) by $\hat{h}_{t+1}$ and due to the linearization of the original observation equation (11). We therefore label $A_{t|t-1}$ and $A_{t|t}$ as approximate MSE matrices. For a given time series $y_1, \ldots, y_T$, the filtering and prediction steps can be carried out recursively for $t = 1, \ldots, T$. The resulting algorithm is known as the extended Kalman filter, see Anderson and Moore (1979) for a more formal derivation. The quasi-loglikelihood function is obtained by inserting the values v_t and F_t, defined below (14), into the loglikelihood (7). We then maximize the quasi-likelihood to obtain estimates for ψ. Estimates of the latent Nelson-Siegel factors, the loading parameter λ_t and the GARCH variance h_t (or g_t) are based on the filtered state estimate $a_{t|t}$.

4 Data and Empirical Findings

For our empirical analysis of yield curves we consider the unsmoothed Fama-Bliss zero-coupon yields dataset, obtained from the CRSP unsmoothed Fama and Bliss (1987) forward rates. We analyze monthly U.S. Treasury yields with maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months over the period from January 1972 to December 2000. This dataset is the same as the one analyzed by Diebold, Rudebusch, and Aruoba (2006) and Diebold and Li (2006) who provide more details on its construction.

[insert Table 1]

Table 1 provides summary statistics for our dataset. For each maturity, we report mean, standard deviation, minimum, maximum and some autocorrelation coefficients. We also present the statistics for proxies of the level, slope and curvature of the yield curve, see the discussion in Section 2.1. The summary statistics reveal that the average yield curve is upward sloping. Volatility decreases by maturity, with the exception of the 6-month being more volatile than the 3-month bill. Yields for all maturities are persistent, most notably for long term bonds. However, with a first-order autocorrelation of 0.970, the 3-month bill is also highly persistent. The level, slope and curvature proxies are persistent but to a lesser extent. The curvature and slope proxies are least persistent given the twelfth-order autocorrelation coefficients of 0.259 and 0.410, respectively.

4.1 DNS: baseline dynamic Nelson-Siegel model

We have been able to estimate the baseline DNS model with parameter estimates that are almost identical to those in Diebold, Rudebusch, and Aruoba (2006, Table 1, p.316). A slight difference stems from our restriction of a stationary VAR process for the factors, see Section 2.2. The factor loadings parameter λ is estimated as 0.0778, with a standard error of 0.00209. The high significance of this estimate confirms that interest rates are informative about λ while small changes in the loadings have a significant effect on the likelihood value.

[insert Table 2]

Table 2 reports sample means and standard deviations of filtered errors. The filtered errors are defined as the difference between the observed yield curve and its filtered estimate, obtained from the Kalman filter. We find that in particular the 3-month rate is difficult to fit: it has the highest mean filtered error. The standard deviations reported in Table 2 indicate that the bonds with intermediate maturity are filtered most accurately.

4.2 DNS–TVL: time-varying factor loadings

To obtain some indication whether the λ parameter varies over time, we consider the baseline model for four equally sized subperiods that cover the full sample. The four estimates of λ for the consecutive subperiods are 0.0397, 0.126, 0.0602 and 0.0695. The corresponding standard errors are sufficiently small to conclude that the four λ estimates are distinct from each other (except for the last two subsamples). This finding provides some evidence that the assumption of constant factor loadings over time does not necessarily hold.

[insert Figure 1]

Next we consider the DNS model by treating the factor loadings parameter λ as a latent factor that is modeled jointly with the other factors by a VAR process, see Section 3.1. We estimate the coefficients of this model and obtain filtered estimates of both the three yield factors and the time-varying λ using the extended Kalman filter discussed in Section 3.3. Panel (A) of Figure 1 presents the filtered estimates of the factor loadings parameter λ. The λ estimates in 1974 are particularly high whereas at the end of the 1970's and the beginning of the 1980's the estimates are rather volatile. Although many changes occur in the early part of the sample, the changes in the late 1990s are also pronounced. Since both slope and curvature of the yield depend on λ, we conclude that sufficient evidence is provided of significant changes in the characteristics of the yield curve over time.

[insert Table 3]

Parameter estimates of the DNS model with λ as a latent factor are discussed in Section 4.4. Here we focus on the fit of the model. Table 2 enables comparisons, for each maturity, between the sample means of the filtered errors for the DNS and DNS–TVL models. For 13 out of the 17 maturities the mean filtered error is lower. This is particularly the case for short maturities. The standard deviations of the filtered errors are lower for 11 out of the 17 maturities. Table 3 reports the performance of the DNS models by presenting values for the loglikelihood, the Akaike Information Criterion (AIC) together with the likelihood-ratio (LR) test for model improvement. When comparing the loglikelihood values between the DNS and DNS-TVL models, the difference of 300 is convincing by any means. This is confirmed by the AIC and LR values. The results therefore provide sufficient evidence of a highly significant improvement in the fit of the DNS–TVL model over its baseline version.

4.3 DNS–GARCH: time-varying volatility

The second modification of the DNS model is to allow for a common time-varying volatility component in the observation disturbances using the GARCH specification discussed in Section 3.2. The details of estimation are discussed in Section 3.3. Panel (A) of Figure 2 presents the filtered estimates of the common volatility. It shows that the common volatility is particularly high in the early years of the 1980's while from the end of the 1980s onwards the volatility is low and rather constant over time. The latter finding may suggest that after the publication of the Nelson and Siegel (1987) paper, their method has become the default of practitioners to price the cross-sections of yields which may have had a dampening effect on volatility. However, low volatility in a prolonged period from the mid-1980's has also been detected for time series of US Inflation, see the discussion in Stock and Watson (2007).

[insert Figure 2]

Table 2 reports the mean of the filtered errors for the model with GARCH and this mean is lower for 15 out of the 17 maturities when compared to those for the baseline DNS model. Only the 72 and 120-month bonds have a higher mean in the DNS–GARCH model. Furthermore, the standard deviations of the filtered errors of the DNS–GARCH model is lower for 12 out of the 17 maturities. In Table 3 we compare loglikelihood and AIC values of the DNS–GARCH model with those of the baseline DNS model. Similarly to the DNS–TVL model, we find a highly significant improvement in the loglikelihood value of the DNS–GARCH model over the baseline model. The likelihood increase and the AIC decrease are even higher than in the case for the DNS–TVL models. It indicates that most gains in describing the yield curve in this dataset are obtained by introducing time-varying volatility.

We also consider the DNS–GARCH model for treating volatility in η_t, the innovations of the factors in (4). In this specification, the GARCH process is loaded onto the level, slope and curvature factors while it is indirectly loaded onto the observed yields via β_t. Empirical support for this specification is weak, the GARCH parameter estimates indicate that the common volatility component is close to a constant while the other parameter estimates are similar to those obtained for the DNS model. The loglikelihood increase of 14.5 reported in Panel (B) of Table 3 is relatively small but it is significant. However, we obtain stronger support for a common GARCH component in the observation disturbances in ε_t. In the latter case, we can test whether the GARCH loadings are linear combinations of the factor loadings via the restriction $\Gamma = \Lambda(\lambda)w$ where w is unknown and needs to be estimated. The resulting loglikelihood increase of 92 compared to the baseline model is significant but moderate when compared to the increase obtained by the unrestricted DNS–GARCH model.

4.4 DNS–TVL–GARCH: time-varying loadings and volatility

Given the encouraging initial results of the last two subsections, we next discuss in more detail the estimation results presented in Table 4 for the DNS–TVL–GARCH model, the DNS model with both time-varying factor loadings and volatility. In Table 2, the means and standard deviations of the filtered errors for the full model specification are given. In comparison with the baseline DNS model, we observe that the filtered error mean is lower for 14 out of 17 maturities. Although this improvement is slightly less than for the DNS–GARCH model, we also have 14 error series that have smaller standard deviations compared to the baseline model. Such improvement has not been obtained by the other DNS models.

The loglikelihood and AIC values reported in Panel (A) of Table 3 for the full model show strong significant improvements compared to the baseline DNS model. When we benchmark the values against models with only time-varying factor loadings or only time-varying volatility, we also obtain significant improvements. We therefore conclude that both model extensions significantly contribute to improvements in the DNS model fit. The GARCH extension provides the most significant improvement.

Panel (A) of Figure 1 presents the filtered λ estimates obtained from the DNS–TVL–GARCH model where λ is treated as a latent factor. The λ estimates are similar to the

DNS–TVL model. Panel (B) presents the loadings for slope and curvature that are obtained using the minimum and maximum value of the estimates of λ. It shows clearly that the loadings can differ significantly over time. The current model specification provides this flexibility. In Panel (A) of Figure 2 the filtered estimate of the common GARCH component is displayed for the DNS–TVL–GARCH model. The volatility estimates are similar to the DNS–GARCH model. However, in the period at the end of the 1980's, the estimates of both λ and the common volatility are different when compared to the single DNS extensions. It is interesting to observe that for this period the filtered λ estimates are lower compared to the DNS–TVL model. The sharp increases in the yields in this period are explained more accurately by a common GARCH component than a time-varying loading parameter λ.

[insert Table 4]

In Table 4 we report a selection of the parameter estimates for the DNS-TVL-GARCH model. We first focus on the estimate of the VAR coefficient matrix Φ for β_t, with the four latent factors, which is reported in Panel (A) of Table 4. When compared to the estimates of the baseline DNS model, reported by Diebold, Rudebusch, and Aruoba (2006, Table 1, p.316), the inclusion of λ as a latent factor mostly affect the dynamics of the slope and curvature. A new empirical finding is the high persistence of the time-varying factor loadings parameter λ. The results also reveal that the curvature factor depends heavily on the factor loadings parameter while, compared to the baseline model, it is less persistent and has a higher variance. The factor loadings parameter λ depends heavily on the (lagged) slope and curvature factors. The estimated variance matrix Σ_η is reported in Panel (B). Although the four innovation series for the factors are all correlated, the strong negative correlation between the curvature factor and the λ factor suggests a substitution effect.

The estimates of the GARCH parameters are presented in Panel (C) of Table 4. Since we estimate all elements in loading vector Γ, the constant in the GARCH specification cannot be identified and is kept at a fixed small value. The remaining estimates for the coefficients γ_1 and γ_2 are significant and they have similar values as the ones for the DNS–GARCH model (not reported here). The estimates of the elements in Γ are presented graphically in Panel (B) of Figure 2. The estimated loadings are displayed by a line-plot against the maturity length. Although the loadings are quite smooth against maturity, it is interesting to find that the maturities of 15 and 18 are relatively less subject to the common GARCH component while the short maturities are most affected by GARCH. When the estimated loadings in Γ are interacted with the GARCH component, we obtain the time-varying volatility for each maturity. Panel (C) in Figure 2 displays the volatility process for a selection of maturities.

[insert Figure 3]

In Figure 3 we compare the filtered latent factors obtained from the DNS–TVL–GARCH model with those from the baseline model and their data-based proxies. The level factors are presented with the 120 month yield, the slopes with the spread of 3 month over 120 month yields and the curvatures with the 24 month yield minus the 3 and 120 month yield. The estimated factors from both models describe the data-based proxies equally well. To highlight the differences in fit of our model extensions, the bottom plots in each panel of Figure 3 present the differences of the factors between the DNS and DNS–TVL–GARCH models. The differences are most pronounced for the slope and curvature factors, particularly in the 1973-1974, 1978-1983 and 1991-1994 periods. It confirms the findings reported in Table 4 from which we learn that the dynamics for slope and curvature have been most affected by our extensions when compared to the baseline DNS model.

4.5 Robustness of empirical results

In this section we study the robustness of our results in three ways: (a) comparison with regression results; (b) model with time-varying splines; (c) results based on a different sample.

(a) Results based on regression. When the VAR specification is discarded in the DNS model, the original Nelson and Siegel (1987) model (2) is obtained and the factors level β_{1t}, slope β_{2t} and curvature β_{3t} can be estimated for each period t using standard regression methods. In case λ is treated as unknown, it can be estimated by nonlinear least squares (NLS), see Diebold and Li (2006). For a given estimate of λ, the factors and the constant variance σ_t^2 in (2) can be estimated by ordinary least squares (OLS).

In Panel (A) of Figure 4, the NLS estimates of λ in the Nelson-Siegel model are displayed (as dots) together with the estimates of factor β_{4t} in the DNS–TVL–GARCH model (solid line) as obtained from the methods described in Section 3.3. The individual NLS estimates are well represented by the estimated fourth factor. In some cases, the λ parameter in the

Nelson-Siegel framework cannot be estimated accurately since the estimation relies on 17 observations only. The analysis based on the DNS–TVL–GARCH model provides estimates of λ using current and past observations. The resulting estimates are therefore based on more data and become more stable as a result. However, the DNS–TVL–GARCH specification is sufficiently flexible to provide an adequate representation of the changes in λ over time.

The dots in the graph of Panel (B) are the OLS estimates of the constant variance σ_t^2 in the Nelson-Siegel model (2) with λ fixed at 0.0609 as in Diebold and Li (2006). The estimated common GARCH component of the DNS–TVL–GARCH model is also presented in this graph (with scale adjustment). It is encouraging that the estimated common GARCH component provides an accurate description of the time-varying volatility in the time series of yields. Deviations between the two estimates can be detected at the end of the 1980's.

(b) Results based on time-varying spline functions. To verify that our empirical findings are not specific to a particular model specification, we next consider λ and the common variance as spline functions of time in the DNS model. In this specification, the model is time-varying and linear, conditional on a set of knot positions (known a-priori in the analysis) and a corresponding set of unknown coefficients. Parameter estimation can be based on the standard Kalman filter methods of Section 2.3. When more knots are chosen, the time-varying smooth functions becomes more flexible. Initially we use spline functions based on five knots which are equally spaced over the time-horizon of the sample.

From the empirical results obtained by a model with a spline function for λ, it has become evident that the factor loadings parameter λ is not constant over time. The LR test statistic indicates that the model with a spline function for λ improves the fit significantly compared to the baseline model.By increasing the number of knots, the time-varying λ estimates come closer to those obtained from the DNS-TVL model and displayed in Panel (A) of Figure 1.

Furthermore, we have considered a spline function for the time-varying common volatility component in the observation disturbances. The positions of the five knots are equally spaced over the time-horizon. The model fit improves significantly for this specification when compared to the baseline model. For this model the estimated volatility is high in the period between 1980 and 1987. Thereafter the variance becomes constant for all maturities. When more knots are introduced, the estimated spline function for the variance gets closer to the estimated GARCH component as displayed in Panel (A) of Figure 2.

(c) Results based on a different sample. From the results presented in Panels (A) of Figures 1 and 2 we have learned that the DNS–TVL–GARCH model particularly captures the variations in both λ and the volatility before 1987. It is therefore interesting to investigate whether the DNS–TVL–GARCH model also provides improvements in model fit for the data-set after 1987. For this purpose, we have re-estimated the baseline DNS model and its extensions for the sub-sample indicated by >1987. The results reported in Panel (A) of Table 3 are reproduced for the sub-sample >1987 in the lower section of Panel (B). We are encouraged by the empirical result that the model fit has increased for the TVL and GARCH extensions of the DNS model based on the >1987 sample. The significant improvements for the GARCH extension of the DNS model are pronounced and most likely due to the volatility changes in the initial period after 1987 and in the middle of the 1990's.

5 Conclusion

The Nelson-Siegel framework provides means for an effective time series analysis of yield data. In this paper we propose two extensions for the dynamic Nelson-Siegel (DNS) model of Diebold, Rudebusch, and Aruoba (2006) where the level, slope and curvature of the yield are treated as dynamic latent factors and modeled by a VAR process. The factor loadings in the DNS model depend on a single parameter that is usually taken as fixed. We show that the factor loading parameter can be estimated accurately from the data. It implies that the data can be highly informative about the factor loadings. Our first contribution concentrates on the question whether the factor loading parameter is constant over time. For this purpose we treat the loading parameter as the fourth latent factor in the DNS model. This nonlinear extension of the DNS model leads to a significant improvement in model fit. Next we turn our attention to the volatility pattern in each of the maturities and we focus on the question whether it is constant over time. For this purpose we introduce a common GARCH volatility component in the DNS model. The common volatility component is multiplied by a loading parameter for each maturity. The GARCH extension of the DNS model provides an even more significant improvement in model fit. The empirical results are obtained for a standard dataset that is analyzed by others in the literature. We have given evidence that our empirical results are robust against alternative model specifications and different sample choices. The general framework of the DNS model allows other modifications for future research.

References

Anderson, B D O & J B Moore (1979). *Optimal Filtering*. Englewood Cliffs:Prentice-Hall.

Ang, A & M Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *J.Monetary Economics 50*, 745–87.

Ansley, C F & R Kohn (1986). A note on reparameterizing a vector autoregressive moving average model to enforce stationarity. *J. Statistical Computation and Simulation 24*, 99–106.

Bianchi, F, H Mumtaz & P Surico (2006). The UK Great Stability: a View from the Term Structure of Interest Rates. Working Paper.

Bollerslev, T (1986). Generalised autoregressive conditional heteroskedasticity. *J. Econometrics 51*, 307–327.

Bowsher, C & R Meeks (2008). High Dimensional Yield Curves: Models and Forecasting. *J. American Statistical Association 103*, forthcoming.

Christensen, J, F X Diebold & S Rudebusch (2007). The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models. Manuscript University of Pennsylvania.

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

Dai, Q & K J Singleton (2000). Specification Analysis of Affine Term Structure Models. *J. Finance 55*(5), 1943–1978.

De Pooter, M (2007). Examining the Nelson-Siegel Class of Term Structure Models. Working Paper.

Diebold, F X & C Li (2006). Forecasting the Term Structure of Government Bond Yields. *J. Econometrics 130*, 337–364.

Diebold, F X, C Li & V Yue (2007). Global Yield Curve Dynamics and Interactions: A Generalized Nelson-Siegel Approach. *J. Econometrics*, forthcoming.

Diebold, F X, S Rudebusch & S Aruoba (2006). The Macroeconomy and the Yield Curve. *J. Econometrics 131*, 309–338.

Doornik, J A (2001). *Object-Oriented Matrix Programming using Ox 3.0* (4th ed.). London: Timberlake Consultants Ltd. See http://www.doornik.com.

Duffee, G (2002). Term Premia and Interest Rate Forecasts in Affine Models. *J. Finance 57*, 405–443.

Duffie, D & R Kan (1996). A Yield-factor Model of Interest Rates. *Journal of Mathematical Finance 6*(2), 379–406.

Durbin, J & S J Koopman (2001). *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press.

Engle, R F, V K Ng & M Rothschild (1990). Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills. *J. Econometrics 45*, 213–238.

Fama, E F & R R Bliss (1987). The Information in Long-Maturity Forward Rates. *American Economic Review 77*, 680–692.

Harvey, A C, E Ruiz & E Sentana (1992). Unobserved component time series models with ARCH disturbances. *J. Econometrics 52*, 129–58.

Heath, D, R Jarrow & A Morton (1992). Bond Pricing and the Term Structure of Interest Rates:a New Methodology for Contingent Claims Valuation. *Econometrica 60*, 77–105.

Hull, J & A White (1990). Pricing interest rate derivative securities. *Rev. Financial Studies 3*, 573–92.

Koopman, S J, N Shephard & J A Doornik (1999). Statistical algorithms for models in state space form using SsfPack 2.2. *Econometrics Journal 2*, 113–66.

Nelson, C & A Siegel (1987). Parsimonious Modelling of Yield Curves. *Journal of Business 60-4*, 473–489.

Stock, J H & M W Watson (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit & Banking 39*, 3–33.

Vasicek, O (1977). An Equilibrium Characterization of the Term Structure. *J. Financial Economics 5*, 177–188.

Yu, W & E Zivot (2007). Forecasting the Term Structures of Treasury and Corporate Yields: Dynamic Nelson-Siegel Models Evaluation. Working Paper.

Table 1: Summary Statistics

The table reports summary statistics for U.S. Treasury yields over the period 1972-2000. We examine monthly data, constructed using the unsmoothed Fama-Bliss method. Maturity is measured in months. For each maturity we show mean, standard deviation (*Std.dev.*), minimum, maximum and three autocorrelation coefficients, 1 month ($\hat{\rho}(1)$), 1 year ($\hat{\rho}(12)$) and 30 months ($\hat{\rho}(30)$).

Summary Statistics for each Maturity							
Maturity	Mean	Std.dev.	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
3	6.851	2.695	2.732	16.020	0.970	0.700	0.319
6	7.079	2.702	2.891	16.481	0.972	0.719	0.355
9	7.201	2.679	2.984	16.394	0.972	0.726	0.378
12	7.302	2.602	3.107	15.822	0.971	0.729	0.394
15	7.408	2.548	3.288	16.043	0.973	0.737	0.415
18	7.481	2.532	3.482	16.229	0.974	0.743	0.431
21	7.544	2.520	3.638	16.177	0.975	0.747	0.442
24	7.558	2.474	3.777	15.650	0.975	0.745	0.450
30	7.647	2.397	4.043	15.397	0.975	0.755	0.470
36	7.724	2.375	4.204	15.765	0.977	0.761	0.480
48	7.861	2.316	4.308	15.821	0.977	0.765	0.499
60	7.933	2.282	4.347	15.005	0.980	0.779	0.514
72	8.047	2.259	4.384	14.979	0.980	0.786	0.524
84	8.079	2.215	4.352	14.975	0.980	0.768	0.526
96	8.142	2.201	4.433	14.936	0.982	0.793	0.535
108	8.176	2.209	4.429	15.018	0.982	0.794	0.540
120(level)	8.143	2.164	4.443	14.925	0.982	0.771	0.532
slope	1.292	1.461	-3.505	4.060	0.929	0.410	-0.099
curvature	0.121	0.720	-1.837	3.169	0.788	0.259	0.076

Table 2: Filtered Errors of Model Extensions

The table reports the filtered errors from the four Nelson-Siegel latent factor models we estimate. The filtered errors are defined as the difference between the observed yield curve and its filtered estimate, obtained from the Kalman filter. The *Baseline* model corresponds to the baseline dynamic Nelson-Siegel latent factor model with constant factor loadings and volatility (DNS). The *Time-Varying Factor Loading* model corresponds to the model with λ added to the state (DNS–TVL). The *Time-Varying Volatility* model corresponds to the model with a common GARCH component for the volatility (DNS–GARCH). The *Both Time-Varying* model corresponds to the model with the factor loadings parameter added to the state and the common GARCH component for volatility (DNS–TVL–GARCH). For each maturity we show mean and standard deviation (*Std.dev.*). We summarize these per model with three statistics: the mean, median and number of maturities for which the absolute value is lower than that of the baseline model (*#Lower*).

Filtered Errors (in basis points)								
	Baseline Model DNS		Time-Varying Factor Loading DNS–TVL		Time-Varying Volatility DNS–GARCH		Both Time-Varying DNS–TVL –GARCH	
Maturity	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
3	-12.63	22.37	-2.87	14.15	-7.61	21.88	-3.29	16.02
6	-1.34	4.87	0.19	1.99	0	0	0.17	0.92
9	0.51	8.13	-0.95	7.54	0.01	9.73	-0.77	7.9
12	1.32	9.89	-0.89	9.46	-0.6	10.53	-1.14	9.61
15	3.72	8.76	1.71	8.29	0.16	5.83	-0.06	5.84
18	3.63	7.22	2.15	6.38	0.72	4.18	0.63	3.73
21	3.26	6.43	2.39	5.82	1.45	6.15	1.87	6.19
24	-1.39	6.33	-1.69	7	-0.87	7.38	-0.66	6.71
30	-2.68	5.98	-2.11	6.35	-1.5	6.42	-1.41	6.05
36	-3.29	6.6	-2.22	6.71	-1.21	5.71	-1.37	5.74
48	-1.83	9.67	-0.52	9.19	1.11	7.6	0.78	7.86
60	-3.29	7.98	-2.3	7.15	-0.79	5.58	-1.46	6.84
72	1.94	9.02	2.41	8.68	2.61	9.01	2.18	9
84	0.68	10.18	0.59	10.6	0.21	11.05	0.58	10.84
96	3.51	9.15	2.9	9.9	1.84	8.83	1.66	8.75
108	4.24	13.5	3.16	13.22	-0.54	7.81	0.48	8.89
120	-1.33	16.34	-2.82	16.43	-4.09	14.7	-3.17	15.82
Mean	-0.29	9.55	-0.05	8.76	-0.54	8.38	-0.29	8.04
Median	0.51	8.76	-0.52	8.29	0	7.6	-0.06	7.86
#Lower			13	11	15	12	14	14

Table 3: Loglikelihood and AIC of Model Extensions

Panel A reports the loglikelihood and Akaike Information Criterion (AIC) for the various model extensions proposed. The *Baseline* model corresponds to the baseline dynamic Nelson-Siegel latent factor model with constant factor loadings and volatility (DNS). The *Time-Varying Factor Loadings* model corresponds to the model with λ added to the state (DNS–TVL). The *Time-Varying Volatility* model corresponds to the model with a common GARCH component for the volatility (DNS–GARCH). The *Time-Varying Loadings and Volatility* model corresponds to the model with the factor loadings parameter added to the state and the common GARCH component for volatility (DNS–TVL–GARCH). In Panel B we report the loglikelihood and AIC for various alternative models and for our extensions estimated only for the period after 1987.

Panel A: Performance of Model Extensions			
	Loglikelihood	AIC	LR-test vs. Baseline
DNS–baseline	3184.6	-6297.1	
DNS–TVL	3484.9	-6875.7	600.6** (0.00)
DNS–GARCH	3657.3	-7204.7	945.6** (0.00)
DNS–TVL–GARCH	3766.8	-7401.7	1164.6** (0.00)

An asterisk (*) denotes significance at the 5% level or less and two asterisks (**) denote significance at the 1% level or less. The probability H_0 is accepted is reported below the test-statistic.

Panel B: Alternative Models and Results for Post-1987 Period			
	Loglikelihood	AIC	LR-test vs. Baseline
Alternative Model Specifications			
DNS–GARCH (in η_t)	3199.1	-6316.2	29.1** (0.00)
DNS–GARCH ($\Gamma = \Lambda(\lambda)w$)	3276.6	-6471.2	184.1** (0.00)
Models Estimated for Post-1987 Period			
DNS–baseline (>1987)	3041.7	-6011.4	
DNS–TVL (>1987)	3213.5	-6333.0	343.6** (0.00)
DNS–GARCH (>1987)	3544.3	-6978.6	1005.2** (0.00)
DNS–TVL–GARCH (>1987)	3668.5	-7205.0	1253.6** (0.00)

An asterisk (*) denotes significance at the 5% level or less and two asterisks (**) denote significance at the 1% level or less. The probability H_0 is accepted is reported below the test-statistic.

Table 4: Estimates of Latent Factors VAR Model and GARCH Process

The table reports the estimates of the vector autoregressive (VAR) model for the latent factors and the GARCH parameter estimates. The results shown correspond to the latent factors of the Nelson-Siegel latent factor model with the time-varying factor loadings parameter added to the state and a common GARCH component for the volatility (DNS–TVL–GARCH). Panel A shows the estimates for the constant vector μ and autoregressive coefficient matrix Φ, Panel B shows the estimates for the covariance matrix Σ_η, Panel C the estimates for the common GARCH process.

Panel A: Constant and Autoregressive Coefficients of VAR

	Level_{t-1}	Slope_{t-1}	Curvature_{t-1}	Loading_{t-1}	Constant (μ)
Level_t ($\beta_{1,t}$)	0.994**	0.0497**	−0.0287*	0.0369	7.82**
	0.00832	0.016	0.0135	0.0437	1.24
Slope_t ($\beta_{2,t}$)	−0.0118	0.931**	0.0149	−0.0165	−1.63**
	0.0133	0.0298	0.0255	0.0629	0.409
Curvature_t ($\beta_{3,t}$)	−0.0308	0.198**	0.658**	0.878**	0.443
	0.0338	0.0646	0.0449	0.17	0.392
Loading_t (λ_t)	0.0179	−0.0555**	0.0734**	0.585**	−2.37**
	0.00948	0.0177	0.0165	0.0639	0.143

An asterisk (*) denotes significance at the 5% level or less and two asterisks (**) denote significance at the 1% level or less. The standard errors are reported below the estimates.

Panel B: Variance Matrix of VAR

	Level_t ($\beta_{1,t}$)	Slope_t ($\beta_{2,t}$)	Curvature_t ($\beta_{3,t}$)	Loading_t (λ_t)
Level_t ($\beta_{1,t}$)	0.0988**	−0.0445**	0.0986**	−0.00393
	0.00899	0.0116	0.0367	0.0157
Slope_t ($\beta_{2,t}$)		0.237**	0.0596	0.0595**
		0.0365	0.0537	0.0203
Curvature_t ($\beta_{3,t}$)			1.6**	−0.302**
			0.22	0.0638
Loading_t (λ_t)				0.188**
				0.0265

An asterisk (*) denotes significance at the 5% level or less and two asterisks (**) denote significance at the 1% level or less. The standard errors are reported below the estimates.

Panel C: GARCH Parameters

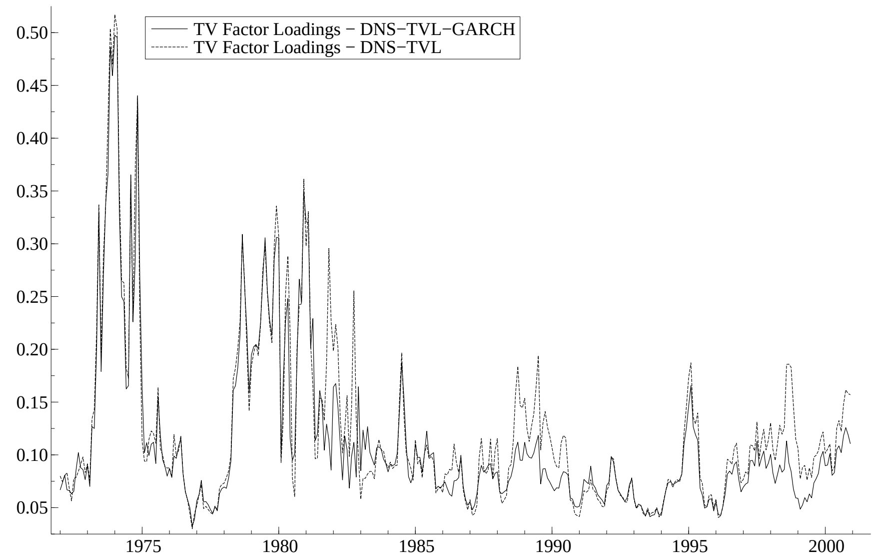
	γ_0	γ_1	γ_2
Estimate	0.0001	0.471**	0.506**
	NA	0.118	0.118

An asterisk (*) denotes significance at the 5% level or less and two asterisks (**) denote significance at the 1% level or less. The standard errors are reported below the estimates.

Figure 1: Time-Varying Factor Loadings Parameter Added to State

In this figure we present the filtered time series of the factor loadings parameter λ and the slope and curvature loadings using the minimum and maximum value of the filtered λ. In Panel (A) we show the filtered time series for both the model with the factor loadings parameter added to the state (DNS–TVL, dotted line) and the model with both the time-varying factor loadings parameter added to the state and a common GARCH component for the volatility (DNS–TVL–GARCH, solid line). In Panel (B) we show the slope and curvature loadings using the minimum and maximum value of the filtered λ for the DNS–TVL–GARCH model.

(A) Filtered Factor Loadings Parameter



(B) Minimum and Maximum loading of Slope and Curvature

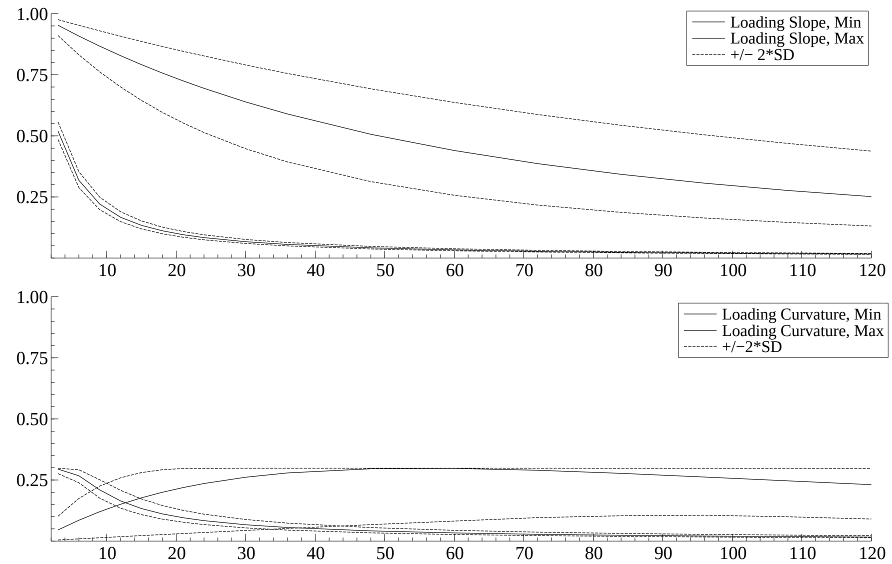
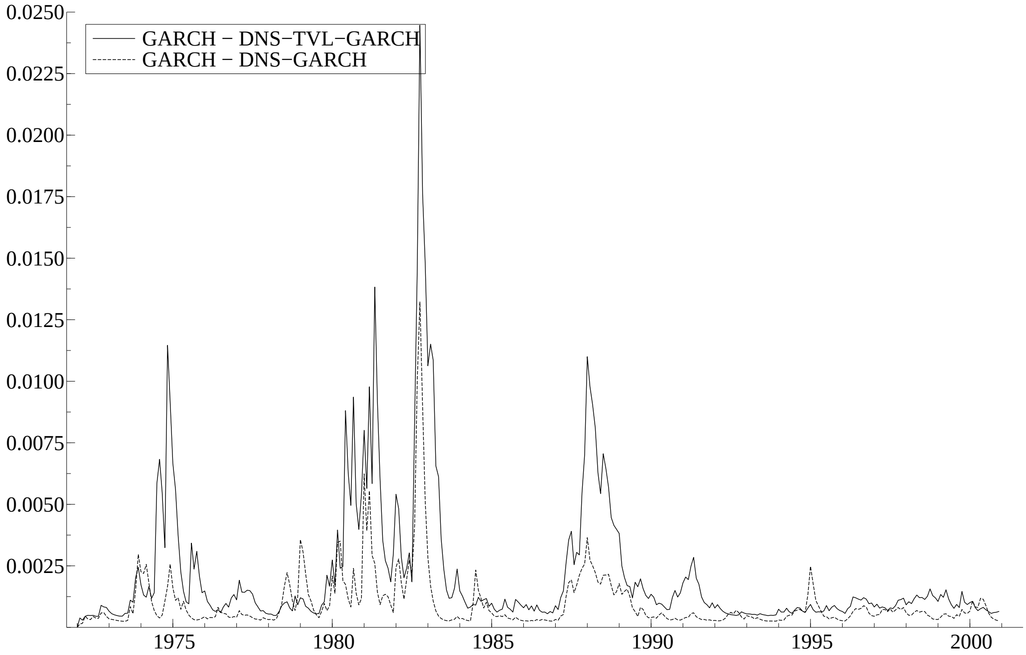


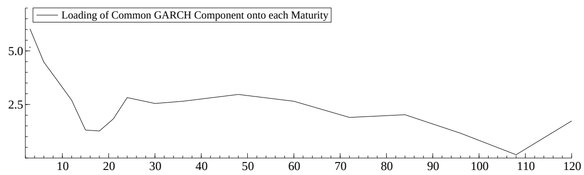
Figure 2: Time-Varying Volatility

In this figure we present the time-varying volatility. In Panel (A) we plot the time varying volatility for both the model with a common GARCH volatility component (DNS–GARCH, dotted line) and the model with both time-varying factor loadings and a common GARCH component for the volatility (DNS–TVL–GARCH, solid line). In Panel (B) we show the loadings, for each maturity, of the common GARCH process in the DNS–TVL–GARCH model. Panel (C) shows the estimated volatility for the DNS–TVL–GARCH model.

(A) Filtered Common GARCH Volatility Component



(B) Loadings of Common GARCH Volatility Component against Maturity



(C) Estimated Volatility for Some Maturities

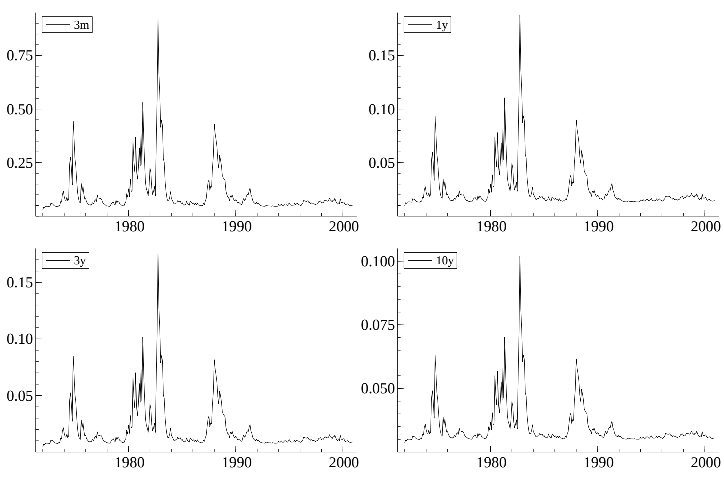
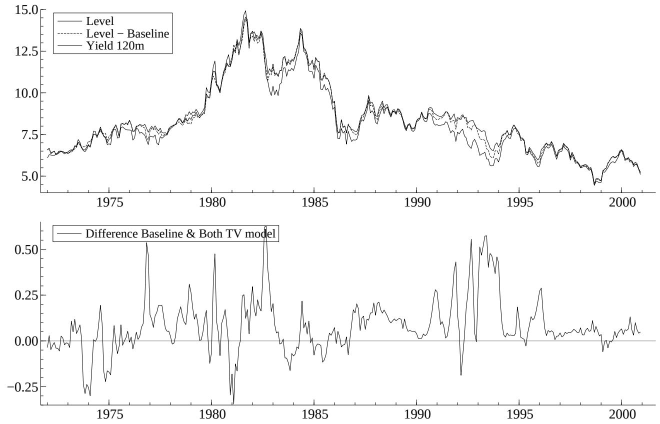


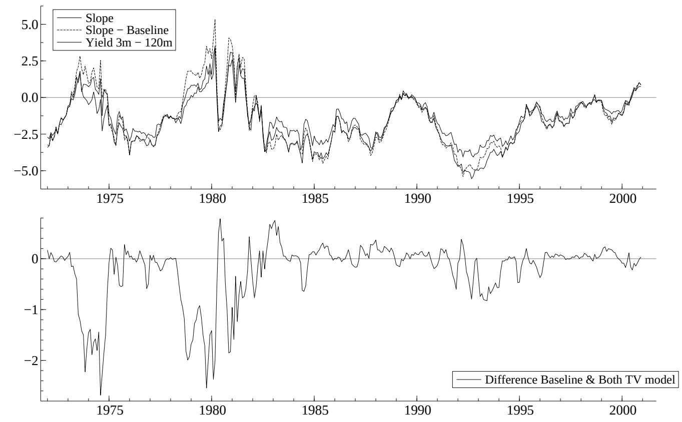
Figure 3: Level, Slope and Curvature

This figure reports the level, slope and curvature as obtained from the Nelson-Siegel latent factor model with both time-varying factor loadings and volatility (DNS–TVL–GARCH). Panels (A), (B) and (C) report the level, slope and curvature respectively together with their proxies from the data. For the level this is the 120 month treasury yield, for slope this is the spread of 3 month over 120 month yields and for curvature this is twice the 24 month yield minus the 3 and 120 month yield. In addition we show the filtered level, slope and curvature for the baseline dynamic Nelson-Siegel model (DNS) and the difference compared to the latent factors from the DNS–TVL–GARCH model (bottom plots in each panel).

(A) Filtered Level



(B) Filtered Slope



(C) Filtered Curvature

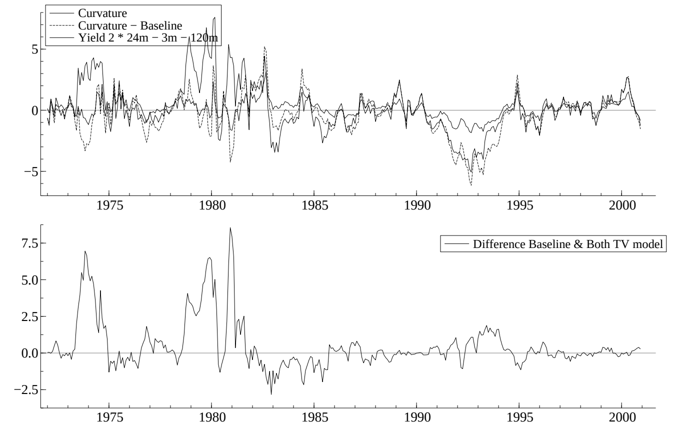
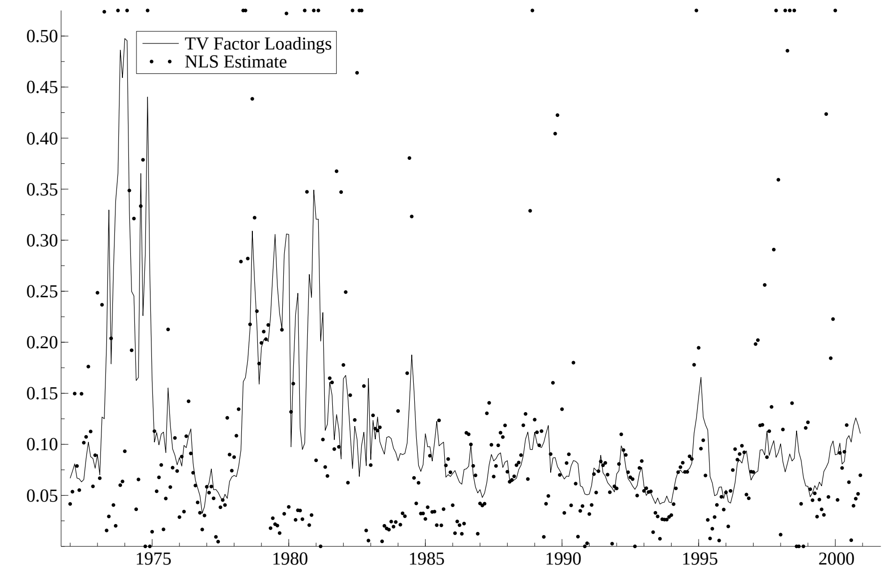


Figure 4: Time-Varying Extensions compared to NLS and OLS Analysis

In this figure we compare the time-varying factor loadings parameter and volatility component from the DNS–TVL–GARCH model to output from the NLS and OLS analysis. We compare the time-varying factor loadings parameter λ to estimates obtained from using NLS. The time-varying common GARCH volatility we compare to the residual variance from the OLS model.

(A) Time-Varying Factor Loadings, compared to NLS



(B) Common Time-Varying Volatility Component, compared to OLS