The Nelson–Siegel Model of the Term Structure of Option Implied Volatility and Volatility Components

BIAO GUO, QIAN HAN* and BIN ZHAO

We develop the Nelson–Siegel model in the context of option-implied volatility term structure and study the time series of volatility components. Three components, corresponding to the level, slope, and curvature of the volatility term structure, can be interpreted as the long-, medium-, and short-term volatilities. The long-term component is persistent and driven by macroeconomic variables, the medium-term by market default risk, and the short-term by financial market conditions. The three-factor Nelson–Siegel model has superior performance in forecasting the volatility term structure, with better out-of-sample forecasts than the popular deterministic implied volatility function and a restricted two-factor model, providing support to the literature of component volatility models. © 2014 Wiley Periodicals, Inc. Jrl Fut Mark 34:788–806, 2014

1. INTRODUCTION

It is recognized that traditional option valuation models such as the Black–Scholes and one-factor stochastic volatility models fail to capture the dynamics of the term structure of implied volatility (see Christoffersen, Heston, & Jacobs, 2009; Park, 2011, for recent discussions). Just as it is accepted in the yield curve literature that one factor is not sufficient to capture the time variation and cross-sectional variation in the term structure of interest rates, the equity option valuation literature acknowledges as critical the use of multifactor models.

Empirical research on the term structure of option-implied volatility (e.g., Byoun, Kwok, & Park, 2003; Diz & Finucane, 1993; Heynen, Kemna, & Vorst, 1994; Mixon, 2007;...
Poteshman, 2001; Stein, 1989) has reached a consensus that long-term volatility reacts differently to volatility shocks than short-term volatility. Therefore, a natural way to extend the one-factor volatility models is to decompose the volatility into long-term and short-term components. Recently, these component volatility models (CVM) (see Bates, 2000; Christoffersen et al., 2009; Christoffersen, Jacobs, Ornthanalai, & Wang, 2008; Park, 2011) have been shown to perform better than one-factor volatility models in modeling the implied volatility term structure.

This article proposes a reduced-form component volatility model to capture the characteristics of option-implied volatility term structure. Our motivation originates from the analogy between the term structure of fixed income derivatives and that of equity options. The interest rate and option-implied volatility term structures are quite similar in many aspects (see Christoffersen et al., 2009; Derman, Kani, & Zou, 1996). Just as each Treasury bond has a corresponding yield to maturity, each traded index option has a corresponding implied volatility. Both the yield curve and implied volatility term structures exhibit a high degree of time variation and cross-sectional variation.

Motivated by this intuition, we develop the Nelson–Siegel model based on a continuous time two-factor volatility model and decompose the implied volatility into three components: the long-term, medium-term, and short-term volatilities. These three volatility components are shown to correspond to the empirical level, slope, and curvature of the term structure of implied volatility. In addition, macroeconomic and financial variables have significant explanatory power for the long-term volatility, suggesting that the long-term volatility is driven by macroeconomic and financial shocks. The short-term component is highly correlated with the VIX index, a popular measure of 1-month volatility expected by market investors. A large proportion of its variations can be explained by unemployment change and stock market growth.

We then model the time series properties of these volatility components and examine the model’s in-sample and out-of-sample performances. For ease of comparison, we use the random walk model as a benchmark and compare the Nelson–Siegel model with an ad hoc term structure model as in Dumas, Fleming, and Whaley (1998) and Pena, Rubio, and Serna (1999). Both in-sample and out-of-sample tests indicate that the simple Nelson–Siegel model performs better than these models, especially for long maturity options. This has particular implications for practitioners because the Nelson–Siegel model is easy and straightforward to implement. All these findings are not sensitive to whether we include the recent financial crisis into the sample or not.

We also show that traditional one-factor volatility models such as the Heston (1993) stochastic volatility model can be reduced to a two-factor Nelson–Siegel model. An encompass test comparing the two-factor and three-factor Nelson–Siegel models suggests that an additional volatility factor is important for option valuation, especially so for pricing long maturity options, consistent with the argument that two state variables are needed for option pricing (Li & Zhang, 2010; Park, 2011).

The article makes several contributions. First, the proposed Nelson–Siegel model is parsimonious and simple to implement. This is in sharp contrast to the estimation complexity involved in existing CVMs in the literature. Second, we shed light on the causes of volatilities. The long-term volatility is driven by real economic growth shocks, whereas short-term volatility is related to monetary policy shocks. This suggests that macroeconomic factors are critical for option valuation. Third, we provide a new perspective to modeling the dynamics of implied volatility surface (IVS). Previous literature models the dynamics of IVS by modeling the time series of extracted principal components (Panigirtzoglous & Skiadopoulos, 2004), higher-order moments (Neumann & Skiadopoulos, 2012), or implied volatility functions (Dumas et al., 1998; Pena
et al., 1999). In comparison, our volatility components approach has better economic interpretations.

The rest of the article is structured as follows. Section 1 develops the Nelson–Siegel model in the option implied volatility context. Section 2 introduces the data and fits the Nelson–Siegel model to examine the time series properties of extracted volatility components. Section 3 models the dynamics of the term structure and examines the model’s out-of-sample performance. Section 4 explores the role of the second volatility component. Section 5 concludes.

2. THE NELSON–SIEGEL MODEL AND THE ANALOGY BETWEEN YIELD CURVE AND IMPLIED VOLATILITY TERM STRUCTURE

The Nelson and Siegel (1987) model and its extension (Diebold & Li, 2006) are widely accepted by industry for forecasting the yield curve due to their simplicity and efficiency. The model assumes that the forward rate curve can be described by

\[ f_t(t) = \beta_{1t} + \beta_{2t} e^{-\lambda t} + \beta_{3t} \lambda t e^{-\lambda t}. \]

If we combine this with the relationship between the yield to maturity and the forward rate, it gives this yield curve function

\[ y_t = \beta_{1t} + \beta_{2t} \frac{1 - e^{-\lambda t}}{\lambda t} + \beta_{3t} \left( \frac{1 - e^{-\lambda t}}{\lambda t} - e^{-\lambda t} \right). \]

This functional form is a convenient and parsimonious three-component exponential approximation, which works very well to predict the yield curve, as shown in Diebold and Li (2006). The three factors \( \beta_{1t}, \beta_{2t}, \) and \( \beta_{3t} \) are of particular interest. The loading on \( \beta_{1t} \) is 1, a constant that does not decay to zero in the limit; hence, \( \beta_{1t} \) may be viewed as a long-term factor. The loading on \( \beta_{2t} \) is \( (1 - e^{-\lambda t})/\lambda t \), a function that starts at 1 but decays monotonically and quickly to 0, and hence may be viewed as a short-term factor. The loading on \( \beta_{3t} \) is \( (1 - e^{-\lambda t})/\lambda t - e^{-\lambda t} \), which starts at 0, increases, and then decays to zero; hence, it may be viewed as a medium-term factor. In addition, Diebold and Li (2006) demonstrate that these three factors may also be interpreted in terms of the yield curve’s level, slope, and curvature, respectively. They further show that this simple model performs better than many other yield curve models in both in-sample fitting and out-of-sample forecasting.

Given the analogy between yield curve and the term structure of option-implied volatility, it is natural to ask whether the nice properties of the model can be carried over to the case of implied volatility, and if it can, then two issues are worth exploring: first, whether the interpretation of the three factors, level, slope, and curvature, is still valid; and second, whether the efficiency of the model in predicting future yield curves can be maintained when the model is applied to forecasting the implied volatility term structure.

2.1. The Nelson–Siegel Model in the Context of Option Valuation

It turns out that the answers to these questions are yes. In this section, we motivate by first developing a two-factor Nelson–Siegel model based on a single volatility model,\(^1\)

\(^1\)Chalamandaris and Tsekrekos (2011) is the first to use the Nelson–Siegel model to approximate the term structure of option-implied volatility. However, they fail to justify the validity of applying this model to the context of option pricing and to link the model factors with volatility components.

\(^2\)Similar derivations are in Stein (1989) and Park (2011).
and then extending it to a standard Nelson–Siegel model by introducing a second volatility factor. Consider a volatility model where the instantaneous volatility $\sigma_t$ evolves according to the following continuous-time mean-reverting AR(1) process:

$$d\sigma_t = -\alpha(\sigma_t - \bar{\sigma}) \, dt + \beta \sigma_t \, dz_t.$$ 

At time $t$, the expectation of volatility as of time $t+j$ is given by (see Theorem 1 in Ait-Sahalia, 1996)

$$E_t(\sigma_{t+j}) = \tilde{\sigma}_t + \rho^j(\sigma_t - \bar{\sigma}),$$

where $\rho = e^{-a} < 1$. Denote by $i_t(T)$ the implied volatility at time $t$ on an option with $T$ remaining until expiration. Then the option-implied volatility is

$$i_t(T) = \frac{1}{T} \int_{j=0}^{T} \left[ \bar{\sigma} + \rho^j(\sigma_t - \bar{\sigma}) \right]dj = \bar{\sigma}_t + \frac{\rho^T - 1}{T \ln \rho} [\sigma_t - \bar{\sigma}].$$

Plugging $\rho = e^{-a}$ back into the above equation and rearranging gives the two-factor Nelson–Siegel model

$$i_t(T) = \bar{\sigma}_t + \frac{1 - e^{-aT}}{aT} (\sigma_t - \bar{\sigma}).$$

Note that for any level of mean reverting speed $\alpha$, the first (second) derivative of implied volatility with respect to maturity is negatively (positively) related to $(\sigma_t - \bar{\sigma})$. Hence when the instantaneous volatility is low (high) relative to its historical level, we shall observe an upward (downward) sloping and concave (convex) term structure. In Figure 1, we plot the actual implied volatility curve and the fitted two-factor implied volatility curve for selected dates. It is quite obvious that the two-factor model can only capture certain shapes of the implied volatility curve but not those with humps. This demonstrates one of the shortcomings of the single volatility model.

Now consider a volatility model where the instantaneous volatility $\sigma_t$ evolves according to the following continuous-time process:

$$d\sigma_t = -\alpha(\sigma_t - \bar{\sigma}) \, dt + \beta \sigma_t \, dz_t,$$

$$d\bar{\sigma}_t = -\kappa(\bar{\sigma}_t - \bar{\bar{\sigma}}) \, dt + \xi \bar{\sigma}_t \, dw_t.$$ 

That is, we assume the instantaneous volatility mean reverts to the medium run volatility $\tilde{\sigma}_t$, which itself follows a mean reverting process with its mean being the long-run volatility $\bar{\sigma}_t$. This volatility decomposition is in the same spirit of the component volatility model in Christoffersen, Jacobs, Ornthanalai, and Wang (2008). At time $t$, the expectation of volatility as of time $t+j$ is then given by

$$E_t(\sigma_{t+j}) = E(\bar{\sigma}_{t+j}) + \rho^j(\sigma_t - E(\bar{\sigma}_{t+j})),
$$

$$E_t(\bar{\sigma}_{t+j}) = (\bar{\sigma}_t) + \tau^j(\bar{\sigma}_t - \bar{\sigma}_t),$$

$^3$Note we do not restrict the long-run mean $\bar{\sigma}_t$ and $\bar{\bar{\sigma}}_t$ to be constants.
where $\rho = e^{-\alpha} < 1, \tau = e^{-\kappa} < 1$. Denote by $i_t(T)$ the implied volatility at time $t$ on an option with $T$ remaining until expiration. Then we have

$$i_t(T) = \frac{1}{T} \int_{j=0}^{T} \left[ \bar{\sigma}_t + \tau' (\sigma_t - \bar{\sigma}_t) + \rho' (\sigma_t - \bar{\sigma}_t) - \bar{\sigma}_t \right] dj = \bar{\sigma}_t + \frac{\rho^T - 1}{T \ln \rho} [\sigma_t - \bar{\sigma}_t] - \rho^T (\bar{\sigma}_t - \bar{\sigma}_t)$$

The last equality is obtained by imposing a restriction on the mean reverting speed parameters $\rho$ and $\tau$ so that the number of parameters is reduced to keep the model parsimonious and avoid the over-parameterization problem as pointed out in the case yield curve fitting (Nelson & Siegel, 1987). Plugging back $\rho = e^{-\alpha}$ into the above equation and rearranging gives

$$i_t(T) = \bar{\sigma}_t + \frac{1 - e^{-\alpha T}}{\alpha T} \left[ \sigma_t - \bar{\sigma}_t \right] - e^{-\alpha T} (\bar{\sigma}_t - \bar{\sigma}_t).$$
Following the argument in Diebold and Li (2006), the above equation can be rewritten as

\[ i_t(T) = \beta_{1t} + \frac{1 - e^{-\alpha_1 t}}{\alpha_1 \tau} \beta_{2t} + \left[ \frac{1 - e^{-\alpha_1 T}}{\alpha T} - e^{-\alpha T} \right] \beta_{3t}, \]

which is in the exact form of the Nelson–Siegel model. In Figure 2, we plot the actual implied volatility curve and the fitted Nelson–Siegel implied volatility curve for selected dates. Compared with Figure 1, the two factor volatility model fits the humps much better because of the extra term of \( \beta_{3t} \) that comes from the introduction of a long-term volatility component.

3. DATA AND PROPERTIES OF VOLATILITY COMPONENTS

In this section, we first describe the sample and then fit the volatility term structure using the Nelson–Siegel model. The time series properties of the model factors are analyzed and their economic implications explained.
3.1. Data

Our sample includes daily S&P 500 index call options from January 3, 2005 to October 29, 2010, providing a total of 1,468 data points. We use the volatility surfaces taken from the Ivy DB OptionMetrics database with ten different time-to-maturities (30, 60, 91, 122, 152, 182, 273, 365, 547, and 730 days) on each trading day. Since not all time-to-maturities are traded on each day, OptionMetrics interpolates the surface to obtain the missing data. We only select those options with delta equal to 0.5 whereas they are the most liquidly traded in the market.

Panel A in Table I reports the mean, maximum, minimum, SD, and autocorrelation for implied volatilities with different time-to-maturities. These descriptive statistics are of interest in themselves. It is clear that the typical implied volatility curve is upward sloping: the mean value becomes larger as time-to-maturity increases. For example, the mean of 30-day volatility is 0.1964 while for 730-day volatility it is 0.2043. Longer maturity volatilities are less volatile and more persistent than shorter ones, indicating the necessity of modeling the long- and short-term volatilities separately. We also compute the empirical level, slope, and curvature of the volatility term structure. The empirical level is defined as the 365-day implied volatility. The slope is the 365-day implied volatility minus the 30-day implied volatility. Finally, the curvature is the 122-day implied volatility minus half of the sum of the 365- and 30-day implied volatilities. As shown in Panel B of Table I, the level is highly persistent with a reasonable value of around 20% for the sample period. The slope is slightly positive, consistent with the observation in Panel A that the volatility term structure is typically upward-sloping for the sample period. The positive curvature implies a concave term structure, on average.

<table>
<thead>
<tr>
<th>Maturity (days)</th>
<th>Mean</th>
<th>Maximum</th>
<th>Minimum</th>
<th>SD</th>
<th>( \hat{\rho}(10) )</th>
<th>( \hat{\rho}(30) )</th>
<th>( \hat{\rho}(60) )</th>
<th>( \hat{\rho}(180) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Implied volatility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.1964</td>
<td>0.7498</td>
<td>0.0832</td>
<td>0.1058</td>
<td>0.9240</td>
<td>0.7910</td>
<td>0.6440</td>
<td>0.2980</td>
</tr>
<tr>
<td>60</td>
<td>0.1978</td>
<td>0.6727</td>
<td>0.0907</td>
<td>0.0972</td>
<td>0.9410</td>
<td>0.8350</td>
<td>0.6950</td>
<td>0.3440</td>
</tr>
<tr>
<td>91</td>
<td>0.1986</td>
<td>0.6066</td>
<td>0.0968</td>
<td>0.0914</td>
<td>0.9510</td>
<td>0.8600</td>
<td>0.7300</td>
<td>0.3780</td>
</tr>
<tr>
<td>122</td>
<td>0.1999</td>
<td>0.5750</td>
<td>0.1022</td>
<td>0.0870</td>
<td>0.9580</td>
<td>0.8730</td>
<td>0.7460</td>
<td>0.3950</td>
</tr>
<tr>
<td>152</td>
<td>0.2006</td>
<td>0.5399</td>
<td>0.1044</td>
<td>0.0830</td>
<td>0.9620</td>
<td>0.8820</td>
<td>0.7620</td>
<td>0.4150</td>
</tr>
<tr>
<td>182</td>
<td>0.2011</td>
<td>0.5044</td>
<td>0.1059</td>
<td>0.0800</td>
<td>0.9650</td>
<td>0.8920</td>
<td>0.7790</td>
<td>0.4310</td>
</tr>
<tr>
<td>273</td>
<td>0.2017</td>
<td>0.4649</td>
<td>0.1096</td>
<td>0.0749</td>
<td>0.9700</td>
<td>0.9050</td>
<td>0.8000</td>
<td>0.4590</td>
</tr>
<tr>
<td>365</td>
<td>0.2022</td>
<td>0.4449</td>
<td>0.1125</td>
<td>0.0722</td>
<td>0.9730</td>
<td>0.9100</td>
<td>0.8070</td>
<td>0.4710</td>
</tr>
<tr>
<td>547</td>
<td>0.2030</td>
<td>0.4019</td>
<td>0.1161</td>
<td>0.0669</td>
<td>0.9760</td>
<td>0.9180</td>
<td>0.8220</td>
<td>0.4910</td>
</tr>
<tr>
<td>730</td>
<td>0.2043</td>
<td>0.3839</td>
<td>0.1174</td>
<td>0.0644</td>
<td>0.9770</td>
<td>0.9200</td>
<td>0.8240</td>
<td>0.4960</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor</th>
<th>Mean</th>
<th>Maximum</th>
<th>Minimum</th>
<th>SD</th>
<th>( \hat{\rho}(10) )</th>
<th>( \hat{\rho}(30) )</th>
<th>( \hat{\rho}(60) )</th>
<th>( \hat{\rho}(180) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel B: Implied volatility curve level, slope, and curvature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td>0.2022</td>
<td>0.4449</td>
<td>0.1125</td>
<td>0.0722</td>
<td>0.9730</td>
<td>0.9100</td>
<td>0.8070</td>
<td>0.4710</td>
</tr>
<tr>
<td>Slope</td>
<td>0.0058</td>
<td>0.3119</td>
<td>-0.0683</td>
<td>0.0465</td>
<td>0.8150</td>
<td>0.5210</td>
<td>0.2810</td>
<td>0.0770</td>
</tr>
<tr>
<td>Curvature</td>
<td>0.0007</td>
<td>0.0294</td>
<td>-0.0552</td>
<td>0.0067</td>
<td>0.4570</td>
<td>-0.0350</td>
<td>-0.0920</td>
<td>-0.0340</td>
</tr>
</tbody>
</table>

Note. This table presents the descriptive statistics of implied volatility, the curve level, slope, and curvature of implied volatility term structure, and estimated factors. The last four columns contain sample autocorrelations at displacements of 10, 30, 60, and 180 days. The sample period is January 3, 2005 to October 29, 2010.
3.2. Fitting the Implied Volatility Curves

We fit the implied volatility curve using the Nelson–Siegel model

\[ y_t = \beta_{1t} + \beta_{2t} \left( \frac{1 - e^{-\lambda_1 t}}{\lambda_1 t} \right) + \beta_{3t} \left( \frac{1 - e^{-\lambda_2 t}}{\lambda_2 t} - e^{-\lambda_3 t} \right). \]

The parameters \( \{\beta_{1t}, \beta_{2t}, \beta_{3t}\} \) are estimated by ordinary least squares (OLS) with \( \lambda_i \) fixed at a pre-specified value of 0.0147.\(^4\) Applying this procedure for each day produces a time series of estimates \( \{\beta_{1t}, \beta_{2t}, \beta_{3t}\} \), as shown in Figure 3.

The vertical lines in Figure 3 mark three significant dates: August 9, 2007 (the start of the global financial crisis), September 15, 2008 (the collapse of Lehman Brothers), and December 31, 2009 (the time of crisis over) respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

3.2. Fitting the Implied Volatility Curves

We fit the implied volatility curve using the Nelson–Siegel model

\[ y_t = \beta_{1t} + \beta_{2t} \left( \frac{1 - e^{-\lambda_1 t}}{\lambda_1 t} \right) + \beta_{3t} \left( \frac{1 - e^{-\lambda_2 t}}{\lambda_2 t} - e^{-\lambda_3 t} \right). \]

The parameters \( \{\beta_{1t}, \beta_{2t}, \beta_{3t}\} \) are estimated by ordinary least squares (OLS) with \( \lambda_i \) fixed at a pre-specified value of 0.0147.\(^4\) Applying this procedure for each day produces a time series of estimates \( \{\beta_{1t}, \beta_{2t}, \beta_{3t}\} \), as shown in Figure 3.

The vertical lines in Figure 3 mark three significant dates: August 9, 2007, the start of the recent global financial crisis, September 15, 2008 when Lehman Brothers collapsed, and December 31, 2009 when the crisis is over. Clearly the three factors were relatively steady before the start of the crisis but became more volatile with the deepening of the crisis. Panel A in Table II presents descriptive statistics for the estimated factors. The mean of \( \beta_{1t} \) is 0.2047, which is quite close to the average volatility during the sample period. The autocorrelations suggest that it is the most persistent factor; therefore, it is natural to view \( \beta_{1t} \) as a long-term volatility component. A further check on the validity of the Nelson–Siegel model shows that the pair-wise correlations between the estimated factors are not large, as shown in Panel B of Table II.

To confirm the stationarity, or otherwise, of these factors, we conduct augmented Dickey–Fuller (ADF) tests under three assumptions: the regression equation contains the

\(^4\)The parameter \( \lambda_3 \) governs the exponential decay rate; small values of \( \lambda_3 \) produce slow decay and can better fit the curve at long maturities, while large values of \( \lambda_3 \) produce fast decay and can better fit the curve at short maturities. \( \lambda_i \) also governs where the loading on \( \beta_{3t} \) achieves its maximum. As a result, we choose \( \beta_{3t} \) value that maximizes the loading on the medium-term (122-day) factor, which gives 0.0147.
constant and time trend terms, it contains only a constant term, and it contains none of these terms. Panel C of Table II displays the results of the ADF tests. All tests fail to reject the null hypothesis that $\hat{b}_1^t$ has a unit root, and all reject the null that $\hat{b}_2^t$ and $\hat{b}_3^t$ have unit roots. The first difference of $\hat{b}_1^t$, however, seems to satisfy the stationarity assumption.

As a visual interpretation of the two factors as long- and short-term volatilities, Figure 4 plots $\hat{b}_1^t$, $\hat{b}_2^t$ along with the VIX index. Compared with the VIX, $\hat{b}_1^t$ moves slowly and smoothly and captures the trend of the volatility very well, verifying that it reflects the long-term volatility. $\hat{b}_2^t$ and VIX mimic each other with a high correlation of 0.8333, consistent with the notion that $\hat{b}_2^t$ reflects the short-term volatility component.

Next, we show that the three factors can be interpreted as the level, slope, and curvature of the volatility term structure, respectively. In the Nelson–Siegel model, an increase in $\hat{b}_{1t}$ raises all implied volatilities equally, as the loading is identical for all time-to-maturities. Thus, it represents the level of the implied volatility curve at a certain point of time. For the slope, recall that it is defined as

$$slope = y(30) - y(365) = 0.6231\hat{b}_{2t} - 0.0156\hat{b}_{3t}.$$  

The second equality uses the optimal value of $\lambda_t = 0.0147$. Clearly, the slope is mainly captured by $\hat{b}_{2t}$. Similarly, the curvature is defined as

$$curvature = y(122) - \frac{1}{2}[y(365) + y(30)] = -0.322\hat{b}_{2t} + 0.1254\hat{b}_{2t}.$$  

### Table II: Descriptive Statistics of the Estimated Factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Mean</th>
<th>Maximum</th>
<th>Minimum</th>
<th>SD</th>
<th>$\hat{b}(10)$</th>
<th>$\hat{b}(30)$</th>
<th>$\hat{b}(60)$</th>
<th>$\hat{b}(180)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Estimated factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{b}_{1t}$</td>
<td>0.2047</td>
<td>0.3496</td>
<td>0.1206</td>
<td>0.0599</td>
<td>0.9760</td>
<td>0.9190</td>
<td>0.8290</td>
<td>0.5180</td>
</tr>
<tr>
<td>$\hat{b}_{2t}$</td>
<td>0.0100</td>
<td>0.5036</td>
<td>-0.1094</td>
<td>0.0762</td>
<td>0.8240</td>
<td>0.5430</td>
<td>0.2980</td>
<td>0.0820</td>
</tr>
<tr>
<td>$\hat{b}_{3t}$</td>
<td>-0.0010</td>
<td>0.2383</td>
<td>-0.2612</td>
<td>0.0418</td>
<td>0.5720</td>
<td>0.2120</td>
<td>0.0590</td>
<td>0.1290</td>
</tr>
<tr>
<td>Panel B: Correlations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{b}_{1t}$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{b}_{2t}$</td>
<td>0.4506</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{b}_{3t}$</td>
<td>0.4561</td>
<td>-0.0492</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel C: Augmented Dickey–Fuller (ADF) test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{b}_{1t}$</td>
<td>-1.9080</td>
<td>0.6498</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First difference of $\hat{b}_{1t}$</td>
<td>-8.3386***</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{b}_{2t}$</td>
<td>-3.4520***</td>
<td>0.0006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{b}_{3t}$</td>
<td>-4.3830***</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Null hypothesis: there is a unit root, \*** Significance at the 1% level.
The dominant factor is $\beta_3$. Figure 5 plots the estimated factors $\{\beta_1t, \beta_2t, \beta_3t\}$ along with the empirical level, slope, and curvature calculated from the data. Each pair tends to move together very closely. In fact, the correlations between the estimated factors and the empirical level, slope, and curvature are very high: $\rho(\beta_1t, \text{level} = 0.9822)$, $\rho(\beta_2t, \text{slope} = 0.9988)$, $\rho(\beta_3t, \text{curvature} = 0.8188)$.

To conclude this section, our findings so far provide strong empirical support for the decomposition of volatility into long-, medium-, and short-term components. The long-term component is persistent and smooth, whereas the short-term component is stationary and more volatile. Any good option pricing model should at least be able to describe the characteristics of these two volatility components in order to capture the dynamics of the term structure of implied volatility. The role of the medium-term component will be examined in more details in a separate section.

3.3. Economic Determinants of Volatility Components

In this section, we explore the economic meanings of extracted volatility components. As shown above, $\hat{\beta}_{1t}$ is related to the long-term volatility component. Engle and Rangel (2008)
and Park (2011) both argue that long-term volatility is associated with macroeconomic shocks. More recently, Bekaert, Hoerova, and Duca (2010) show that the VIX commoves with monetary policy shocks. Whereas the short-term volatility component $\beta^2_{it}$ is highly correlated with the VIX, we conjecture that it is also related to monetary policy variables. To test these hypotheses, we run the following regressions:

$$
\beta_{it} = \alpha + \gamma \beta_{i,t-1} + \psi X_t + \epsilon_t,
$$

where $i = 1, 2, 3$, $X_t$ is a vector of control variables including: 10-year treasury constant maturity rate, new privately owned housing units started, industrial production index, Moody A/AA corporate bond yield, payroll employment, producer price index, cyclically adjusted price earnings ratio, and the S&P 500 monthly index. Note that before running regressions we detrend all growth variables such as industrial production growth, producer price index, and payroll employment growth by taking the logarithmic differences over last month. Table III reports the pairwise correlations of these variables. It is obvious that some variables are highly correlated. To deal with this potential multicollinearity issue, stepwise regressions are used such that only relevant variables are included.

To further consolidate our model, a series of diagnostic analyses are performed on the residuals for each regression. Heteroskedasticity tests suggest that there are no heteroskedasticity, and the LM tests indicate no autocorrelation in the residuals. Finally, unreported stationary tests on the residuals show that the residual series is stationary, excluding the possibility of spurious regressions.

FIGURE 5

Model-based factors versus empirical level, slope, and curvature. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
Panel A of Table IV presents regression results for the long-term volatility component. The \( R^2 \) of the regression is 0.941, and the \( F \)-statistic is 254.22, both suggesting a good fit of model. Four macroeconomic and financial variables are significant in explaining 95% of the total variations in long-term volatility. All of these variables, the 10-year treasury constant maturity rate, new privately owned housing units started, payroll employment growth, and cyclically adjusted price earnings ratio have a negative effect on long-term volatility. This is consistent with the general observation that market volatility tends to be high when the economy underperforms.

Panel B of Table IV presents the results for the short-term volatility component. The payroll employment growth and stock market growth are significant, suggesting that the short-term volatility component is driven by both macroeconomic and financial market conditions. The negative sign of stock market return reflects the well-documented leverage effect between index returns and volatility.

Panel C of Table IV presents the results for the medium-term volatility component. Market default spread, as approximated by the difference between the 10-year treasury rate and Moody’s AAA corporate bond yield, is significant, indicating that the medium-term component is intimately related to market default risk. It is interesting that the sign of stock market return is now positive, further suggesting that the medium-term volatility component is very different in nature from the short-term volatility component.

4. MODELING AND FORECASTING THE IMPLIED VOLATILITY TERM STRUCTURE

One advantage of using the Nelson–Siegel model for forecasting the volatility of term structure is that the forecasting of an entire curve is reduced to the forecasting of several factors. In this section, we first model the Nelson–Siegel factors as ARMA processes individually, followed by modeling them jointly with a VAR model. Their forecasting
performances are then compared with the random walk benchmark and an ad hoc term structure model. In Section 4, we also compare it with the single volatility model. We estimate and forecast the three factors using data from March 3, 2005 to July 26, 2010, with the forecast period being from July 27, 2010 to October 29, 2010.

4.1. ARMA and VAR Models for Forecasting

In our first attempt, we model each Nelson–Siegel factor as an ARMA process. The general form of an ARMA($p$, $q$) model is

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}.$$ 

The implied volatility forecasts based on the underlying ARMA factor specifications are then given by

$$\hat{\gamma}_{t+h/t} = \hat{\beta}_{1,t+h/t} + \hat{\beta}_{2,t+h/t} \frac{1 - e^{-\lambda \tau}}{\lambda \tau} + \hat{\beta}_{3,t+h/t} \left( \frac{1 - e^{-\lambda \tau}}{\lambda \tau} - e^{-\lambda \tau} \right),$$

### TABLE IV
Economic Determinants of Volatility Components

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>SE</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Long-term volatility Component</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_1$</td>
<td>-0.0146</td>
<td>0.0051</td>
<td>-2.8357</td>
</tr>
<tr>
<td>$X_2$</td>
<td>-0.0000</td>
<td>0.0000</td>
<td>-2.599</td>
</tr>
<tr>
<td>$X_5$</td>
<td>-4.2128</td>
<td>1.5367</td>
<td>-2.7414</td>
</tr>
<tr>
<td>$X_7$</td>
<td>-0.0029</td>
<td>0.0013</td>
<td>-2.3092</td>
</tr>
<tr>
<td>$C$</td>
<td>0.2679</td>
<td>0.0449</td>
<td>5.9693</td>
</tr>
<tr>
<td>$Y_1(-1)$</td>
<td>0.3818</td>
<td>0.0933</td>
<td>4.0911</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.9532</td>
<td>256.73</td>
<td></td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.9495</td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

Panel B: Short-term volatility component

| $X_5$       | -5.7034 | 2.2771      | -2.5047 | 0.0148   |
| $X_8$       | -0.7671 | 0.0646      | -11.880 | 0.0000   |
| $C$         | -0.0088 | 0.0039      | -2.2475 | 0.0280   |
| $Y_2(-1)$   | 0.4349  | 0.0761      | 5.7128  | 0.0000   |
| R-squared   | 0.8190  | 98.0208     |         |
| Adj. R-squared | 0.8106     | 0.0000       |

Panel C: Medium-term volatility component

| $X_4 - X_1$ | 0.0387  | 0.0047      | 8.3105  | 0.0000   |
| $X_8$       | 0.1860  | 0.0462      | 4.0296  | 0.0001   |
| $C$         | -0.0598 | 0.0070      | -8.5637 | 0.0000   |
| R-squared   | 0.5256  | 36.57       |         |
| Adj. R-squared | 0.5112     | 0.0000       |

Note. The definitions of the variables are as follows: $Y_1$, the fitted value of $\hat{\beta}_{11}$; $Y_2$, the fitted value of $\hat{\beta}_{21}$; $X_1$, 10-year treasury constant maturity rate; $X_2$, new privately owned housing units started; $X_3$, industrial production growth; $X_4$, Moody AAA corporate bond yield; $X_5$, payroll employment growth; $X_6$, producer price index inflation; $X_7$, cyclically adjusted price earnings ratio; $X_8$, S&P 500 monthly index growth; $Y_1(-1)$, lag of the fitted value of $\hat{\beta}_{11}$, $Y_2(-1)$, lag of the fitted value of $\hat{\beta}_{21}$. The sources for our macroeconomic and financial fundamental variables are the Board of Governors of the Federal Reserve System, the U.S. Department of Commerce: Census Bureau, the U.S. Department of Labor: Bureau of Labor Statistics, Automatic Data Processing, Inc., the Stock Market Data Used in “Irrational Exuberance,” Princeton University Press, and Yahoo! Finance.
where

\[ \hat{\beta}_{1,t+h/t} = \hat{\gamma}_i + \hat{\phi}_{1,1} \hat{\beta}_{1,t-1} + \hat{\phi}_{1,2} \hat{\beta}_{1,t-2} + \cdots + \hat{\phi}_{p,1} \hat{\beta}_{1,t-p} + \hat{\varepsilon}_t + \hat{\theta}_{1,1} \hat{e}_{t-1} + \cdots + \hat{\theta}_{q,1} \hat{e}_{t-q}. \]

We use the Box–Jenkins method to help decide the order of the AR and MA parts. The ACF and PACF for \( \hat{\beta}_{1,t} \), indicate that for \( \hat{\beta}_{1,t} \), an ARMA(0, 1) model is appropriate, whereas for \( \hat{\beta}_{2,t} \) and \( \hat{\beta}_{3,t} \), ARMA(3, 0) and ARMA(3, 2) models are appropriate, respectively. Diagnostic analysis is conducted to ensure the sufficiency of the model.

Table V reports the estimated coefficients and statistics. All coefficients are significant. Notice that the R-squared for estimating \( \hat{\beta}_{1,t} \) is very small (2.4%) compared with those for \( \hat{\beta}_{2,t} \) and \( \hat{\beta}_{3,t} \). This is not surprising because, as shown earlier, \( \hat{\beta}_{1,t} \) is much more persistent than the other two factors and it is first-order differenced.

We also model the factors based on a multivariate VAR(3) process with the order being selected by the Schwartz criterion. A stability check shows that the VAR(3) model is stable. Whereas there are too many parameters in the model, to save space we do not report the VAR(3) model estimates; however, they are all significant and available subject to request.

### 4.2. Out-of-Sample Forecasting Performance of the Three-Factor Model

A good approximation of implied volatility curve dynamics should not only fit well in-sample, but also forecast well out-of-sample. As we assume that the implied volatility curve depends only on \( \{\hat{\beta}_{1,t}, \hat{\beta}_{2,t}, \hat{\beta}_{3,t}\} \), forecasting the implied volatility curve is equivalent to forecasting these factors.

We estimate and forecast the three factors using the ARMA and VAR models as in the previous section. For comparison purposes, we consider two ad hoc models as the following:

\[ y_t = a_t + b_t \tau + c_t \tau^2, \]

\[ y_{t+h|t} = y_t(\tau), \]
where $t$ is the maturity date. The first ad hoc model is similar to the deterministic implied volatility function (IVF) of Dumas et al. (1998) and Pena et al. (1999), whereas the second model is just a random walk which serves as a natural benchmark. Volatility forecast errors at $t + h$ are defined as $y_{t+h}(\tau) - y_{t+h/t}(\tau)$. Descriptive statistics of the absolute forecast errors of 30, 60, 182, 365, 547, and 730 days maturity are reported in Panel A of Table VI. All errors are divided by the errors in the random walk model.

Several observations can be made from Table VI. First, option-implied volatility seems to be predictable at least 1-day ahead for short-term maturity options when transaction cost is not taken account of. This is consistent with the findings in Chalamandaris and Tsekrekos (2011) that in the currency options market the Nelson–Siegel model is helpful in predicting the dynamics of IVS. Second, the longer maturity the option has, the better the Nelson–Siegel model performs relative to the IVF ad hoc model. This demonstrates the advantage of volatility decomposition for modeling the term structure of implied volatility. For example, for the 30-day maturity options, the ratio of average absolute forecast error for the deterministic volatility function model and that for the Nelson–Siegel model is 1.2, whereas the ratio increases to 1.4 for the 182-day maturity options, to 1.6 for the 365-day maturity options, then to 1.6 for the 547-day and to 1.7 for the 730-day maturity options.

### 4.3. Prediction Out of Sample: Long-Term Options

Another criterion for a satisfactory implied volatility curve model is that it is able to predict implied volatility beyond the maturity range of the sample used to fit it. In this section, we predict the implied volatilities of 547- and 730-day maturity options using all other options left in the sample. We use the same methods as previous to fit the implied volatility curves. Table VII displays the prediction results. Both the mean absolute error (MAE) and mean absolute percentage error (MAPE) demonstrate very small pricing errors for 547- and 730-day maturity options (Figure 6).
The Nelson-Siegel Model of Option Implied Volatility

**TABLE VII**
MAE and MAPE for Implied Volatility Prediction

<table>
<thead>
<tr>
<th>Maturity (day)</th>
<th>MAE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>547</td>
<td>0.0025</td>
<td>0.0122</td>
</tr>
<tr>
<td>730</td>
<td>0.0040</td>
<td>0.0206</td>
</tr>
</tbody>
</table>

*Note.* MAE, mean absolute error; MAPE, mean absolute percentage error.

**FIGURE 6**
Actual and predicted implied volatilities of 547-day and 730-day maturity options. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
4.4. Period Without Financial Crisis

As a robustness check, we consider the time period that excludes the recent global financial crisis to see whether the model still maintains good forecasting performance. The sample period starts from January 3, 2005, and ends with May 31, 2007, including 606 trading days. The forecasting period begins in March 1, 2007, and extends through May 31, 2007.

We use the same ARMA method as before to model the time series \( \{\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}\} \). The ACF and PACF indicate that, for \( \hat{\beta}_{1t} \), an ARMA(2, 0) model seems appropriate, whereas for \( \hat{\beta}_{2t} \) and \( \hat{\beta}_{3t} \), ARMA(3, 0) and ARMA(2, 0) models are sufficient, respectively. A multivariate VAR(2) process is also selected with the order being determined by the H–Q criterion.

Descriptive statistics of the absolute forecast errors of various maturity options are reported in Panel B of Table VI, which again shows that the Nelson–Siegel models forecast future implied volatility very well and produce smaller forecast errors for longer maturity options relative to the IVF method.

5. THE ROLE OF MEDIUM-TERM VOLATILITY COMPONENT

We have shown that popular single volatility models such as the Heston stochastic volatility model corresponds to a two-factor Nelson–Siegel model:

\[
y_t = \beta_{1t} + \beta_{2t} \left( \frac{1 - e^{-\lambda t \tau}}{\lambda t \tau} \right),
\]

where volatility is decomposed into two parts. In the standard Nelson–Siegel model, we decompose the implied volatility into three parts: long-term, short-term, and medium-term. A natural question arises: is it really necessary to include the medium-term component in the option valuation models? To address this question, we repeat the previous fitting and forecasting exercises and report the statistical details of forecasting errors in Table VIII. Note that the errors are divided by the errors in the three-factor model.

We find that the three-factor Nelson–Siegel model outperforms the two-factor model for all the maturities. To corroborate our findings, an ENC-NEW test, as proposed in Clark and McCracken (2001, 2005), is applied to compare the equal forecast accuracy between the two-factor and three-factor models. Using the two-factor model as the null and the three-factor Nelson–Siegel model as the alternative, Table VIII shows that we cannot reject the null hypothesis for short maturity options; however, for those long maturity options with

<table>
<thead>
<tr>
<th>Maturity (days)</th>
<th>ENC-NEW</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.1489 [0.4408]</td>
<td>1.0589</td>
</tr>
<tr>
<td>60</td>
<td>0.1321 [0.4474]</td>
<td>1.0192</td>
</tr>
<tr>
<td>182</td>
<td>−1.4130 [0.9212]</td>
<td>1.0733</td>
</tr>
<tr>
<td>365</td>
<td>1.9051 [0.0284]</td>
<td>1.0047</td>
</tr>
<tr>
<td>547</td>
<td>5.5644 [0.0000]</td>
<td>1.0511</td>
</tr>
<tr>
<td>730</td>
<td>6.6206 [0.0000]</td>
<td>1.1272</td>
</tr>
</tbody>
</table>

Note. Mean absolute forecast errors are divided by the errors for a three-factor Nelson–Siegel model. P values of the ENC-NEW tests are reported in the brackets.
more than 365-day maturities, the three-factor model easily beats the two-factor model. This suggests that the component volatility models are essential for modeling the implied volatility term structure, especially for long maturity options.

6. CONCLUSION

This article is the first empirical study of the volatility components extracted from the term structure of the S&P 500 index option implied volatility. Inspired by the similarity between the modeling of the interest rate term structure and the modeling of implied volatility term structure, we develop the Nelson–Siegel model in the context of options and study the time series of three volatility components in the model. We show that these components, corresponding to the level, slope, and curvature of the volatility term structure, can be interpreted as the long-, medium-, and short-term volatilities. The long-term component is persistent, and the short-term component is highly correlated with the VIX index. We further demonstrate that macroeconomic and financial variables help explain these volatility components. The long-term component is driven by macroeconomic variables, the medium-term by market default risk, and the short-term by financial market conditions. These are revealing because after decades of research existing literature have failed to link these variables to option pricing. We also show that the Nelson–Siegel model has superior performance in forecasting the volatility term structure, compared with the popular implied volatility function method. Finally, we demonstrate that the three-factor Nelson–Siegel model is better in out-of-sample prediction than a two-factor model, providing support to the burgeoning literature of component volatility models.

Future work may include using a state-space model and Kalman filter to extract the volatility components instead of the two-stage method as in the article. A more rigorous comparison of the Nelson–Siegel model with other reduced form volatility term structure models may also be warranted.

REFERENCES


